Medical Image Registration: Statistical Models of Performance in Relation to the Statistical Characteristics of the Image Data

by

Michael Daniel Ketcha

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
July, 2020

Abstract

For image-guided interventions, the imaging task often pertains to registering preoperative and intraoperative images within a common coordinate system. While the accuracy of the registration is directly tied to the accuracy of targeting in the intervention (and presumably the success of the medical outcome), there is relatively little quantitative understanding of the fundamental factors that govern image registration accuracy.

A statistical framework is presented that relates models of image noise and spatial resolution to the task of registration, giving theoretical limits on registration accuracy and providing guidance for the selection of image acquisition and post-processing parameters. The framework is further shown to model the confounding influence of soft-tissue deformation in rigid image registration — accurately predicting the reduction in registration accuracy and revealing similarity metrics that are robust against such effects. Furthermore, the framework is shown to provide conceptual guidance in the development of a novel CT-to-radiograph registration method that accounts for deformation.

The work also examines a learning-based method for deformable registration to investigate how the statistical characteristics of the training data affect the ability of the model to generalize to test data with differing statistical characteristics. The analysis provides insight on the benefits of statistically diverse training data in generalizability of a neural network and is further applied to the development of a learning-based MR-to-CT synthesis method.

Overall, the work yields a quantitative approach to theoretically and experimentally relate the accuracy of image registration to the statistical characteristics of the image data, providing a rigorous guide to the development of new registration methods.

Thesis Committee Members

Jeffrey H. Siewerdsen, Ph.D. (Primary Advisor)
Professor, Department of Biomedical Engineering
Johns Hopkins University

Jerry L. Prince, Ph.D.
Professor, Department of Electrical and Computer Engineering
Johns Hopkins University

J. Tilak Ratnanather, D.Phil.
Associate Research Professor, Department of Biomedical Engineering
Johns Hopkins University

Joshua T. Vogelstein, Ph.D.
Assistant Professor, Department of Biomedical Engineering
Johns Hopkins University

Dedication:

For Mom and Dad

Acknowledgements

I first want to thank my advisor Jeff Siewerdsen who has been an amazing source of inspiration and knowledge since the day I interviewed to join the I-STAR lab. His dedication to science and to the success of his students is incomparable, and I am grateful to have had the pleasure to learn so much from him on every aspect of the scientific process. Along with him I would also like to thank my committee (Jerry Prince, Tilak Ratnanather, and Josh Vogelstein) for their valuable feedback on my dissertation.

I must also thank all of my lab mates in the I-STAR lab who have been a tremendous team that has greatly contributed to my work, from direct involvement to small chats over espresso — Ali, Tharindu, Alex, Wojtek, Jen, Ja, Hao, Joseph, Matt, Sarah, Qian, Runze, Pengwei, Esme, Niral, Sophia, Rohan, Prasad, and Craig. I particularly want to thank Tharindu and Ali who have been great mentors since my first days of grad school and have provided much advice, knowledge, and code over the years.

I also want to acknowledge the research and clinical collaborators who have directly and indirectly aided my research through providing data, methods, feedback, and innovation — Web Stayman and the AIAI lab, Junghoon Lee (Radiation Oncology), Nick Theodore (Neurosurgery), Stan Anderson (Neurosurgery), Clifford Weiss (Interventional Radiology), and Nafi Aygun (Radiology).

Finally, I want to thank my family and the friends that I have made over my years in Baltimore. My parents (Dan and Marcia), to whom this thesis is dedicated, who have provided unending support and love in everything I have set out to do. My brother Marc who I always

have looked up to. My friends and roommates who have provided so many moments that I look back on with joy, from climbing up a mountain to chatting on a porch. And Shanna, who was has added so much to my life since the day I met her.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Clinical Background and Significance

1.1.1 Medical Imaging Modalities

Medical images are widely used in diagnostic and interventional procedures for their ability to non-invasively provide accurate visualization and localization of internal anatomy. Imaging modalities that are commonly used in such settings include x-ray projections, computed tomography (CT), and magnetic resonance (MR), among others — with example images of the spine shown in Fig. 1.1. The modalities are distinct in terms of the physics of image formation, contrast mechanism, and scanner technology, and each has its own advantages and disadvantages. Some essential principles of the imaging modalities that will be used and referred to throughout this thesis are summarized below.

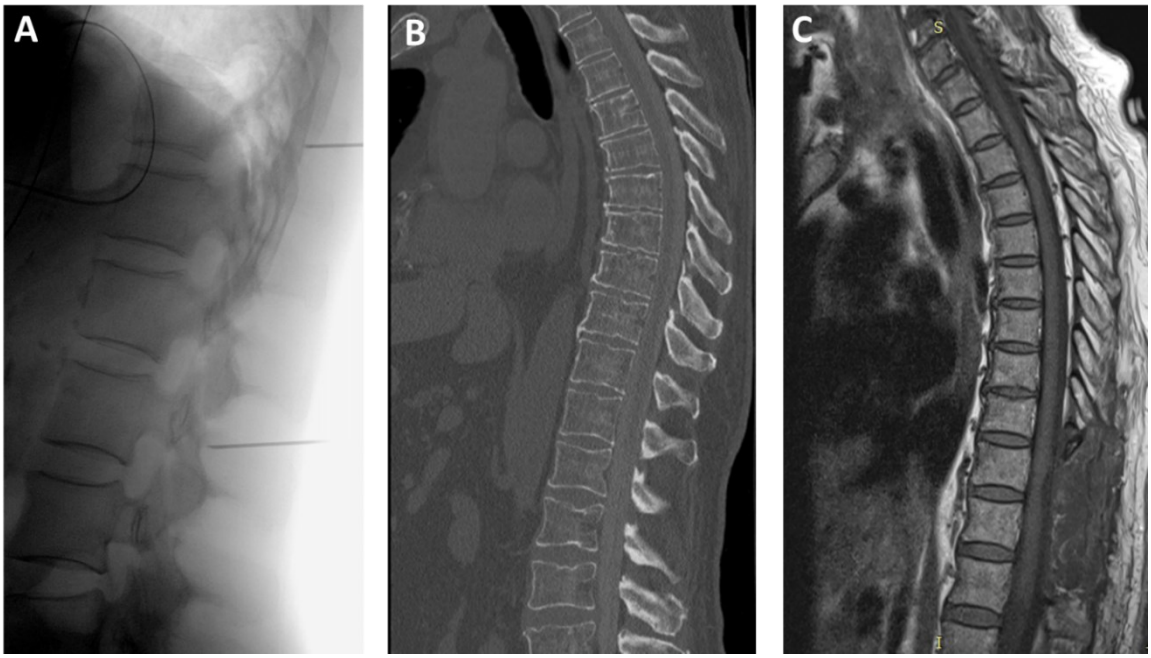


Figure 1.1: Example images of the spine from various modalities including (A) x-ray projection (lateral view; notice a pair of needles used to localize vertebral levels during intervention), (B) CT (sagittal view showing bone and soft tissue structures), and (C) T1-weighted MR (sagittal view; note the neoplastic disruption of tissues posterior to the spinal canal in the lower right of the image and the associated stenosis / compression of the spinal cord).

In x-ray projection imaging, an x-ray source emits high energy (up to 150 keV) photons in a collimated beam toward the patient and a detector. Some of the photons are attenuated or scattered in the patient according to the linear attenuation of materials in the body, and the transmitted beam is incident upon an area detector (e.g., a flat-panel detector, FPD) to generate the projection image. Because structures such as bone and metal have stronger attenuation to x-rays, detector signal in the region of such structures is low. As illustrated in Fig. 1.1A, bones and metal appear accordingly dark in the image, while air and soft tissues appear bright. Of course, the displayed image intensities may be inverted (resulting in bright bones and black air) and nonlinearly transformed to improve the displayed dynamic range of image intensities. X-ray projection imaging is used in screening (e.g., mammography), diagnosis (e.g., fractures and pneumonia), and surgical intervention (e.g., localizing target anatomy and instrumentation).

X-ray projection systems are relatively inexpensive, have fast image acquisition times, and may be portable. X-ray projection images may also be acquired continuously using a pulsed x-ray beam and detector with real-time readout — referred to as fluoroscopy (cf., a single-shot projection image, referred to as a radiograph). The primary limitations of x-ray projection imaging come from the projective nature of the modality, which results in overlapping image content (e.g., lungs and vertebrae in the superior region of Fig. 1.1A) such that the contrast of soft-tissue structures is limited, and the 3D position of structures can be challenged. Multiple images from different angles are often acquired to allow the user to better perceive 3D relationships.

In CT imaging, projection x-ray data are acquired from multiple angles around the patient to reconstruct a 3D volume of the anatomy. As such, it is based on the same contrast mechanism (attenuation coefficient) as x-ray projection imaging, but the tomographic nature of CT image reconstruction provides dramatic improvement in soft-tissue contrast (as shown in Fig. 1.1B) and accurate 3D localization of structures. CT images are commonly used in diagnosis (e.g., detection of tumors, hemorrhages, and fractures) and as preoperative images for purposes of surgical planning. Multidetector CT (MDCT) scanners typically require a dedicated room in which the scanner is fixed to the floor (not portable), although portable systems have also been developed.

Cone-beam CT (CBCT) operates according to similar principles as MDCT with the primary difference coming from the increased area (viz., the longitudinal extent) of the detector (usually a FPD). The volumetric beam (cf., relatively narrow fan-beam collimation in MDCT) permits acquisition of a volumetric image in a single rotation about the patient. The volumetric beam also results in higher levels of x-ray scatter reaching the detector, which reduces soft-

tissue contrast and causes artifacts (shading and streaks) in CBCT image reconstructions. However, due to the smaller footprint, portability, and ability to acquire radiography, fluoroscopy, and/or CBCT, such systems have become fairly common in interventional settings. Disadvantages of both CT and CBCT include biological hazards associated with ionizing radiation to both the patient and interventional staff.

MR imaging utilizes a strong magnetic field paired with radio frequency (RF) impulses to alter the net magnetic moment of nuclei (primarily protons in hydrogen atoms) in the patient along the direction of the primary magnetic field. An RF pulse is used to excite the magnetic moments to a net transverse magnetization, and as the RF pulse is turned off, the magnetic moments return to alignment with the direction of the primary magnetic field and emit an RF signal that is measured by receiver coils surrounding the patient. By varying the pulse sequence and the measurement protocol, various contrast mechanisms are available — some of the most common being T1 (Fig. 1.1C) and T2 weighted images related to the longitudinal and transverse relaxation times of nuclear spins, respectively. In a given imaging session, multiple sequences may be acquired to exhibit various contrast mechanisms. Typically, MR imaging is utilized for clear differentiation of soft-tissue structures — e.g., in preoperative planning of spine or brain surgery, identifying cancerous lesions, torn ligaments, and ischemic stroke. Disadvantages of MR imaging include the relatively high cost of the system, long scan times, and lower spatial resolution. MR scanners also tend to have a large footprint and strong restrictions for MR-compatible materials in the nearby environment.

Other imaging technologies that are worth mentioning (though not directly addressed in the body of this dissertation) include ultrasound, nuclear medicine, and optical imaging. Ultrasound images are acquired by emitting pressure waves into the patient and measuring the

echo signal that is returned to yield an image in which the intensity is related to the acoustic reflectance (impedance) of tissue interfaces in the body. Ultrasound systems tend to carry moderate cost, excellent portability, and real-time image acquisition; however, ultrasound systems carry a high degree of user variability and are not applicable to every anatomical site (e.g., challenged in imaging through bone or gas). [1] Nuclear medicine includes various forms of emission computed tomography (ECT) — e.g., positron emission tomography (PET) and single photon emission computed tomography (SPECT) — in which a molecular compound tagged with a radioactive tracer is introduced in the body. [2] The molecular compound distributes in the body and undergoes particular physiological interactions (e.g., glucose metabolism), and the radioactivity emitted by the tracer is detected by sensor arrays about the patient to localize the site of emission. Optical imaging applied to medical diagnosis and interventional guidance comes in many forms, the most common being optical microscopy and endoscopy. Optical imaging is rich with opportunities to sense a wide range of contrast mechanisms and physiological processes through the use of fluorescent agents [3], [4].

1.1.2 Image-Guided Interventions

Image-guided intervention refers generally to a medical procedure in which images of the patient are utilized to provide anatomical visualization in order to guide delivery of a local therapy. Preoperative images (e.g., CT or MRI) are commonly used to help define the interventional plan and provide surrounding anatomical context. Many scenarios require imaging during the procedure to provide an up-to-date visualization of the patient anatomy, surgical tools, and implanted devices. Image guidance is particularly common in the fields of image-guided radiotherapy (IGRT), interventional radiology (e.g., embolization and other

catheter-based procedures), and some surgical procedures (e.g., orthopedics, neurosurgery, and head and neck surgery). In each of these scenarios, multiple images are acquired — before the procedure ("preoperative," for purposes of diagnosis and planning), during the intervention ("intraoperative," to guide therapeutic delivery), and after the procedure ("post-operative," to confirm the surgical product and check against complications.

For example, in IGRT careful attention is paid to define segmentations of normal and target tissues within a preoperative CT (or MR) image. During the procedure, CBCT images are acquired to determine the current position of the target. In the case of surgery, image guidance can be vital to minimally invasive procedures in which surgery is delivered through small incisions and endoscopic surgical tools. Therefore, imaging systems (e.g., x-ray fluoroscopy and CBCT) provide an updated view on the progress of the procedure. For example, in deep brain stimulation (DBS) electrode placement, careful attention is taken to define the target placement of the electrodes within preoperative MR images, and an intraoperative CBCT scan is acquired to register information from the MR images to the patient position and a robotic arm that aids in electrode placement. Furthermore, multiple radiographs and a post-instrumentation CBCT may be acquired to verify placement of the electrodes while the patient is still in the operating room. Another example is spine surgery, where minimally invasive techniques are utilized to place screws down the pedicles of the vertebrae as a basis for spinal fixation. In this case, while surgeons use preoperative CT or MR images to define the surgical approach, multiple radiographic images and/or CBCT are acquired during the procedure to ensure that the screws are placed properly.

An important step in many image-guided interventions is the transformation of information defined in the preoperative image (e.g., surgical/therapeutic targets locations) to

the intraoperative image. In spine surgery, for example, the target and surgical plan (e.g., pedicle screw trajectory) may be defined in the preoperative CT, and the location of such information within the patient at the time of surgery are given by intraoperative CBCT. One method to relate the preoperative and intraoperative contexts is by geometric alignment through image registration. Image registration comprises a framework as illustrated in Fig. 1.2 in which a similarity metric is defined between the two images — one referred to as "moving" (I_1) and the other as "fixed" (I_2) — and a geometric transformation is applied to the moving image in order to optimize (e.g., maximize) the similarity metric. The transformation that maximizes the similarity metric aligns the moving and fixed images and can be used to relate information defined in the preoperative (moving) image to the intraoperative (fixed) context. In this thesis, an important consideration is the accuracy of this geometric transformation, particularly with respect to the statistical characteristics of the images being registered.

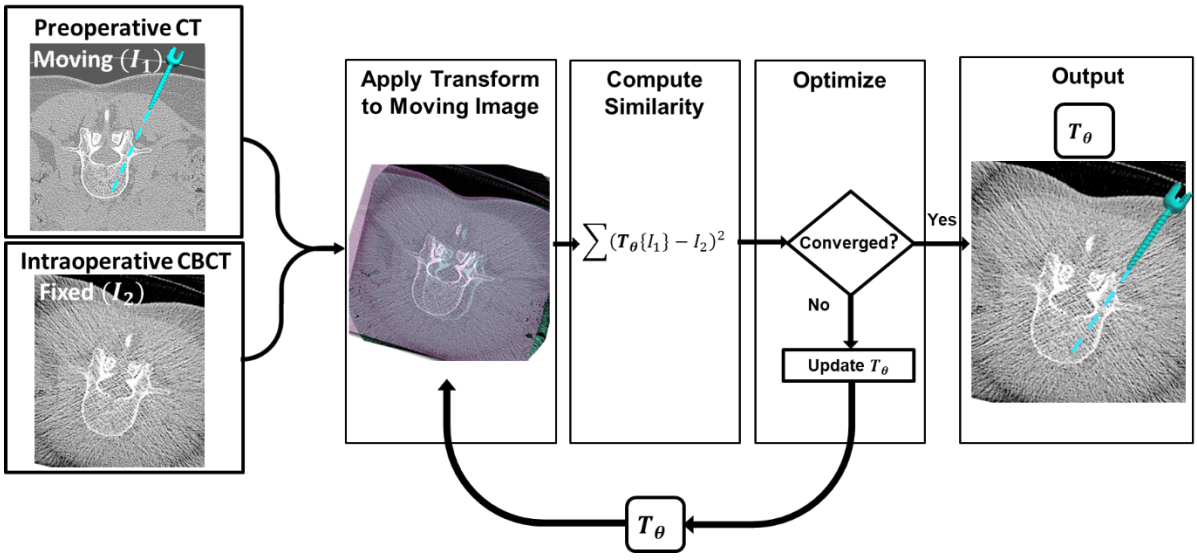


Figure 1.2: Simple flowchart depiction of image registration for image-guided interventions. In this illustration, CT-to-CBCT registration is achieved by maximizing a similarity metric (sum of squared difference in pixel values), resulting in a transformation, theta, that relates the coordinate frames of the preoperative CT and intraoperative CBCT. Information defined in the preoperative CT (e.g., pedicle screw trajectory) can thereby be transformed to the intraoperative CBCT.

1.2 Statistical Descriptors of Imaging Performance

1.2.1 Descriptors of Image Noise and Spatial Resolution

Quantifying imaging performance is an important step in understanding how to optimize an imaging system, image acquisition protocols, and post-processing parameters. Common quantitative descriptors of imaging performance relate to the first- and second-order statistics of the imaging systems, which in turn relate to the spatial resolution and noise, respectively.

The spatial resolution of an imaging system relates to the smallest size at which an image feature is discernable. Pixel/voxel size is often referenced in relation to spatial resolution, but there are many factors that govern this aspect of imaging performance (and a system with small pixel size can still suffer poor spatial resolution, or blur). Quantitative metrics of spatial resolution are often obtained from measurement using a specifically designed imaging phantom. A common method is to use a line-pair phantom that contains multiple sets of parallel lines — each with a characteristic spacing between the lines — and the smallest discernable line pair can be reported as a measure of limiting spatial resolution (typically with units of line pairs/cm). While this is a useful metric, it can be subjective in selecting the smallest discernable line pair and only provides a discrete number of possible spatial resolution levels.

A more complete description of spatial resolution is given by the point-spread function (PSF), which describes the response of the imaging system to a point impulse. Of course, measuring the PSF comes with its own challenges as it, by definition, relies on imaging an infinitesimally small point object — e.g., a pinhole projection in radiography or a fine speck in CT.

Alternatives to the PSF are often obtained by imaging lines or edges to measure the line-spread function (LSF), which is a practical means to determine spatial resolution characteristics in a particular direction. In both cases, such point- and line-spread functions provide an objective characterization of the spatial resolution, which is sometimes simplified to a scalar description by reporting either the full width at half the maximum (FWHM) of the spread function or by fitting a Gaussian and reporting the characteristic width (σ). Alternatively, the magnitude of the Fourier transform of the PSF (or derivative of the LSF) yields the modulation transfer function (MTF) which describes the response of the system as a function of spatial frequency.

Image noise is another important consideration in analysis of image quality. Image noise generally refers to stochastic fluctuations in the image. While the origin of image noise is specific to the imaging modality, noise can be contributed by stochasticity in the signal reaching the detector (e.g., variations in the number of photons reaching the detector in x-ray or camera imaging, referred to as quantum noise) and stochasticity in the electronic readout within the imaging system (referred to as detector readout noise). A basic quantitative descriptor of noise is given by the standard deviation (σ) in image intensity within a region of interest (ROI) that is otherwise uniform. In simplest terms, the visibility of structures in an image is related to the noise in comparison to the contrast of structures of interest — i.e., the contrast-to-noise ratio (CNR):

$$CNR = \frac{\mu_1 - \mu_2}{\sigma} \tag{3}$$

where the difference in mean image intensity (μ) between two features is divided by the noise. However, the contrast and noise do not describe spatial correlations that may be present in the

image. For example, note that CNR can be arbitrarily improved by blurring the image (e.g., by application of a smoothing filter), which reduces σ; of course, doing so would not be expected to generally improve image quality. Therefore, a spatial-frequency-dependent characterization of the noise can be obtained in terms of the noise-power spectrum (NPS), [5] which characterizes not only the noise magnitude but also the spatial correlations.

Factors that contribute to the MTF and NPS are often predictable under assumptions of linearity and shift invariance and can be modelled by analyzing each step in the imaging chain using cascaded linear systems analysis. For instance, sophisticated models have been developed x-ray and CT imaging that capture the spatial resolution and noise transfer characteristics at each stage in the image acquisition process from the interaction of x-rays to the scintillator, to the detection, integration, and discrete sampling of the photons. [6]–[10] Such models can serve as a basis for optimizing each step in the imaging chain with respect to the theoretical ideal imaging system through information-theoretic descriptors such as noise-equivalent quanta (NEQ) and detective quantum efficiency (DQE).

1.2.2 Task-Based Evaluation of Imaging Performance

To paraphrase Barrett [11]: *an image is acquired for a purpose, and performance should be evaluated with respect to that purpose*. While the metrics discussed in Sec. 1.2.1 are useful in quantifying the first- and second-order statistical characteristics of the imaging system (i.e., spatial resolution and noise, respectively), they do not directly describe the performance of the image with respect to a particular imaging task. For example, a common imaging task is that of detection, where a user seeks to detect a particular stimulus (e.g., a lesion) within the image.

As a toy example, Fig. 1.3 (left) shows two stimuli of different spatial extent that a user may want to detect. On the right are three noisy images, each containing the stimulus and with the same noise magnitude, σ, but with increasing noise correlation. Arguably, the stimulus in the top row is most easily detected in the right-most image, while the stimulus in the bottom row is most easily detected in the left-most image. Interestingly, this example suggests that detection performance is best when the spatial-frequency content of the stimulus is mismatched to that of the noise. Clearly, one cannot solely rely on the MTF or NPS alone to understand how an image performs with respect to a detection task, and it is important to also include information about the contrast and spatial-frequency content of the object to be detected.

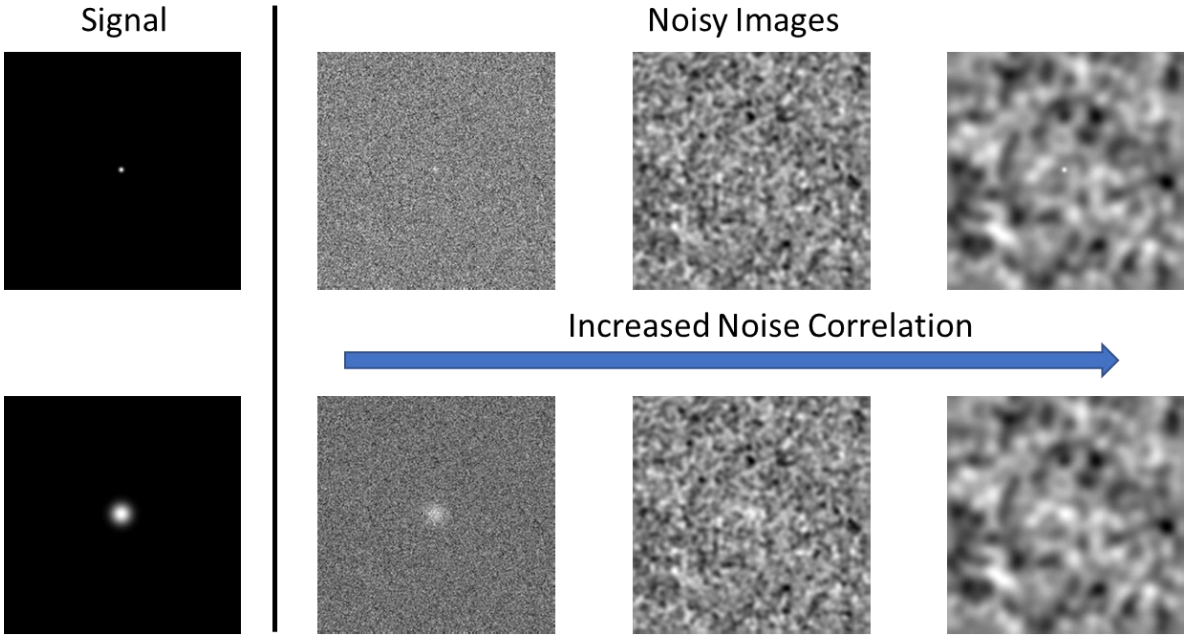


Figure 1.3: Noiseless signals (left) are contaminated by three forms of image noise, each with increased noise correlation but the same noise magnitude (standard deviation, σ).

Due to this phenomenon, a considerable body of ongoing research seeks to establish reliable relationships between image quality and detection tasks [11], [12]. The core of this body of research draws from the field of signal detection theory (SDT), where a variety of

figures of merit have been derived to quantify the ability to differentiate overlapping signals. The first figure of merit that we will discuss is the Hotelling observer model, which defines the detectability index (d') as:

$$d'^2 = \iint \frac{MTF^2(f_x, f_y) \cdot W_{Task}^2(f_x, f_y)}{NPS(f_x, f_y)} df_x df_y \tag{1.1}$$

where f_x and f_y are the spatial frequencies corresponding to a 2D image and W_{Task}^2 is the signal power spectrum of the object that the observer is tasked with detecting, sometimes referred to as the "task function." In this form, we see a direct trade-off as a function of spatial frequency between the image of the object (W_{Task}^2 filtered by the MTF^2 of the imaging system) and the NPS, indicating a benefit when the frequency content of the noise is different from that of the signal (as observed in Fig. 1.2). The model is also referred to as a prewhitening observer model, as it can be shown that the optimal method for detection under this model is to use the prewhitening matched filter which removes the spatial correlation in the noise.

A model that may better correspond to human observer performance is the non-prewhitening observer described by:

$$d'^2 = \frac{\left[\iint MTF^2(f_x, f_y) \cdot W_{Task}^2(f_x, f_y) df_x df_y\right]^2}{\iint NPS(f_x, f_y) \cdot MTF^2(f_x, f_y) \cdot W_{Task}^2(f_x, f_y)\, df_x df_y} \tag{1.2}$$

which describes the performance of a non-prewhitening matched filter (NPWMF). Equation (1.2) contains many similarities to Eq. (1.1); however, the numerator (signal) and denominator (noise) terms are now integrated separately. To gain basic intuition for the NPWMF form, note that the numerator is related to the correlation of the noiseless signal with itself, and the denominator is related to the correlation of the noiseless signal with the noise-only image.

Observer models such as these have been used to design new imaging systems [13], [14], determine optimal imaging techniques [15], and substantiate claims of low-dose imaging performance [16], [17]. Variations on each have been developed to better account for the human observer — e.g., by noting that humans are relatively insensitive to the DC-frequency [18]. Furthermore, a large body of work has shown that not only does stochastic image noise negatively impact the ability to detect an object, but background anatomy can also act as a confounding influence in a detection task — e.g., breast tissue parenchyma when attempting to detect lesions in a mammogram. Interestingly, the effect can be incorporated in a fairly straightforward manner into the observer models of Eqs. (1.1) and (1.2) by modifying the noise term by addition of the power spectrum of the background anatomy (S), i.e., $NPS \rightarrow NPS + S$. The formulation therefore treats the background anatomy as a source of "noise." The power spectrum of background anatomy is often modeled as a power-law distribution:

$$S(f_x, f_y) = \frac{\alpha}{f^{\beta}} \tag{1.3}$$

where $f = \sqrt{f_x^2 + f_y^2}$, α is a scalar related to signal intensity and contrast, and the power-law β scalar determines the low-frequency extent of the background content. Depending on the type of anatomy and imaging modality, β has been shown to be typically in the range of 2–4 [19]–[22], where larger values describe increasingly clumpier background texture.

1.3 Principles of Image Registration

1.3.1 Motion Model

As discussed in Sec. 1.1.2, image registration aims to determine the transformation that geometrically aligns two images, which in the case of image-guided interventions often pertains to registering a preoperative image to an intraoperative image. One of the first considerations for an image registration method is the motion model, which is selected based on type of deformation that is expected to have occurred between the two images. Generally, this is divided among rigid and deformable registration models. For alignment of bone structures in the images, a rigid motion model may be appropriate, whereas a deformable motion model may be necessary for aligning soft-tissue structures. For convenience, we will constrain the discussion below to 3D-3D (volume-to-volume) image registration, recognizing that the principles apply to 2D-2D (e.g., slice-to-slice registration), 3D-2D (e.g., volume-to-slice or volume-to-projection registration), and other registration scenarios.

1.3.1.1 Rigid Registration

Rigid registration refers to global geometric transformations that comprise only translations and/or rotations of the image content. The 3D image transformation itself is described by a 6 degree of freedom (DOF) parameterization of motion, with 3 translation and 3 rotation parameters. Rigid registration is applied in scenarios in which no (or minor) deformation is expected (e.g., the patient has not moved between scans), where the structure

of interest is rigid (e.g., bone anatomy) or where a rigid initialization is performed as input to a subsequent deformable registration.

1.3.1.2 Deformable Registration

Deformable registration relates to recovering spatially varying motion that occurred within the image content, which is often related to deformation in soft-tissue anatomy — e.g., lungs, abdominal structures, and brain parenchyma. Motion models for deformable registration can generally be categorized between parametric and non-parametric methods for defining the deformation.

Parametric methods for image registration are often used for their simplicity and relatively fast computational runtime. A relatively simple form of parametric deformation is the affine transform, which pertains to a 12 DOF parameterization, including shear and scale as well as translation and rotation. Affine transformations are relatively limited in the deformations that can be described and are therefore often used simply to initialize deformable registration methods that are better able to model the anatomical deformation.

One such deformable registration method is a cubic B-spline registration [often called free-form deformation (FFD)], in which a grid of control points is defined in the moving image, and the deformation model is parameterized by the spatial shift (translation) of each of the control points. Cubic B-spline interpolation among the displaced control point positions is utilized to determine the deformation field for the entire image. Coarser or finer levels of motion are enforced by changing the number of control points — where fewer control points can prevent non-realistic deformation but are less able to recover complex motion patterns.

Non-parametric methods for deformable registration relate to those that directly estimate a motion profile for each voxel, thereby generating dense displacement fields on the moving image. Smoothness constraints are incorporated on the deformation field to ensure realistic deformation patterns and improve robustness to image noise during optimization. These constraints often follow from physical models of deformation, such as elastic [23], fluid [24], [25], and diffusion models [26].

Among the most popular of these methods is a variant of the Demons algorithm [27] which approximates a viscoelastic model by incorporating smoothing on both the update fields and the displacement fields to allow recovery of large deformation while staying robust image noise. However, even with these physical constraints on smoothness, there is no guarantee that the topology of the space is preserved, which can lead to overlapping tissue regions and non-invertible transformations. As such, diffeomorphic methods have been utilized, which model the displacement through either time-dependent [28] or stationary velocity [29], [30] that guarantee preservation of topology and smooth invertibility of the transform.

1.3.2 Similarity Metrics

Similarity metrics are functions computed between the images being registered and are maximized (or possibly minimized, depending on the metric) at correct geometric alignment of the images. Below, some common similarity metrics are described, in each case assuming the images to be scalar intensity images.

One subset of similarity metrics for intra-modality registration (i.e., registration of two images of the same modality, where correspondence in image intensities can be safely assumed) is L2-norm similarity metrics, such as the sum of squared differences (SSD):

$$SSD(I_1, I_2|\mathcal{T}_{\boldsymbol{\theta}}) = \sum_i \left(\mathcal{T}_{\boldsymbol{\theta}}\{I_1\}(i) - I_2(i)\right)^2 \tag{1.4}$$

where the $\mathcal{T}_{\boldsymbol{\theta}}\{I_1\}$ is the transformed moving image under the transformation parameters $\boldsymbol{\theta}$, and the metric is summed over each of the i voxels in the images. It is also common for the metric to be normalized by the number of voxels to yield the mean-squared difference similarity (MSD).

Correlation-based similarity metrics are also commonly used in intra-modality registration scenarios. The most basic form is the cross-correlation (CC):

$$CC(I_1, I_2|\mathcal{T}_{\boldsymbol{\theta}}) = \sum_i \mathcal{T}_{\boldsymbol{\theta}}\{I_1\}(i) I_2(i) \tag{1.5}$$

However, a more common form is the normalized cross-correlation (NCC):

$$NCC(I_1, I_2|\mathcal{T}_{\boldsymbol{\theta}}) = \frac{\sum_i (\mathcal{T}_{\boldsymbol{\theta}}\{I_1\}(i) - \overline{\mathcal{T}_{\boldsymbol{\theta}}\{I_1\}})(I_2(i) - \overline{I_2})}{\sqrt{\sum_i \left(\mathcal{T}_{\boldsymbol{\theta}}\{I_1\}(i) - \overline{\mathcal{T}_{\boldsymbol{\theta}}\{I_1\}}\right)^2 \sum_i (I_2(i) - \overline{I_2})^2}} \tag{1.6}$$

where $\overline{I_2}$ is the mean over I_2 and $\overline{\mathcal{T}_{\boldsymbol{\theta}}\{I_1\}}$ is the mean over $\mathcal{T}_{\boldsymbol{\theta}}\{I_1\}$ in the region that overlaps with $\overline{I_2}$. Due to the normalization over the overlapping window, the metric is robust to many false optima and linear differences between the images. In deformable registration, NCC is computed over local windows at each voxel in I_2 (rather than globally over the entire image) and therefore carries an increased computational burden.

Another correlation-based metric is gradient correlation (GC), which is useful when the high-fidelity features coincide with the high gradient features (such as bone):

$$GC(I_1, I_2|\mathcal{T}_{\boldsymbol{\theta}}) = CC\left(\frac{\partial I_1}{\partial x}, \frac{\partial I_2}{\partial x}\middle|\mathcal{T}_{\boldsymbol{\theta}}\right) + CC\left(\frac{\partial I_1}{\partial y}, \frac{\partial I_2}{\partial y}\middle|\mathcal{T}_{\boldsymbol{\theta}}\right) + CC\left(\frac{\partial I_1}{\partial z}, \frac{\partial I_2}{\partial z}\middle|\mathcal{T}_{\boldsymbol{\theta}}\right) \tag{1.7}$$

Note that GC is computed as the sum over the cross correlation among the partial derivative of images in each spatial direction.

The metrics described above rely on direct correspondence between the image intensities of the two images and are therefore most applicable to intra-modality image registration. In the case of multi-modality image registration (e.g., registration of a CT to and MR image), common similarity metrics are based on mutual information (MI). The basic principles of MI come from information theory and Shannon entropy, [31] where the metric describes the shared information between two signals. In its most basic form, MI is defined by creating histograms for each image with k bins and the probabilities (p) for each bin are computed:

$$MI(I_1, I_2|\mathcal{T}_{\boldsymbol{\theta}}) = \sum_{m}^{k}\sum_{n}^{k} p(m,n)\log\frac{p(m,n)}{p_{\mathcal{T}_{\boldsymbol{\theta}}\{I_1\}}(m)p_{I_2}(n)} \tag{1.8}$$

where $p(m,n)$ is the joint probabilities between the two histograms. The form described above, however, is not differentiable due to the discrete histogram binning; therefore, the histogram step is commonly replaced by the use of Parzen windows (e.g., B-splines) to generate a smooth distribution over which MI may be computed [32]. Furthermore, normalized variants such as normalized mutual information (NMI) [33] and entropy correlation coefficient (ECC) [34] have been proposed to reduce the sensitivity to the extent of image overlap.

1.3.3 Optimization

To achieve registration, the geometric transformation parameters $\boldsymbol{\theta}$ (vector of length M) must be optimized according to the similarity metric. While some similarity metrics are designed to be minimized (e.g., SSD) and others maximized (e.g., NCC, MI), for convenience of notation we will assume all metrics will be minimized to achieve a loss function $l(\boldsymbol{\theta})$, typically enforced by simply negating those metrics that should be maximized. For problems in image registration, we are generally confined to local optimizers which require proper initialization to reach the solution. The optimization methods used can generally be divided into two categories: derivative-based and derivative-free optimization. Derivative-based metrics are more commonly used in registration as they tend to provide faster convergence, particularly when $\boldsymbol{\theta}$ is large (e.g., deformable registration); however, they require that the similarity metric is differentiable with respect to the transformation parameters.

A commonly used method for derivative-based optimization is gradient descent. It is a first-order optimization method described by updating $\boldsymbol{\theta}$ according to the gradient of the loss function $\nabla_{\boldsymbol{\theta}} l(\cdot)$ at the current estimate $(\boldsymbol{\theta}^k)$**:**

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - \alpha \nabla_{\boldsymbol{\theta}} l(\boldsymbol{\theta}^k) \tag{1.9}$$

where α is the learning rate which dictates the step size at each iteration. A common draw-back of gradient descent is that it can be slow to optimize and sensitive to local optima (particularly if the learning rate is small). To overcome this, momentum-based methods have

been shown to improve runtime and "push through" local optima while still utilizing only the first-order gradient:

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^{k} - \alpha \nabla_{\boldsymbol{\theta}} l(\boldsymbol{\theta}^{k}) + \beta(\boldsymbol{\theta}^{k} - \boldsymbol{\theta}^{k-1}) \tag{1.10}$$

which updates the gradient descent method by further including a scalar multiple (β) of the previous step. Careful selection of β must be taken, as the method may tend to overshoot the solution if it is too high.

Second-order derivative based metrics can further reduce the number of iterations needed to converge if the loss function is twice-differentiable. A common approach in this scenario is the Newton method, where the update is provided by:

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^{k} - \alpha \left(\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} l(\boldsymbol{\theta}^{k}) \right)^{-1} \nabla_{\boldsymbol{\theta}} l(\boldsymbol{\theta}^{k}) \tag{1.11}$$

where $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} l(\boldsymbol{\theta}^{k})$ is the Hessian matrix of the loss function with respect to $\boldsymbol{\theta}$. While the number of iterations may be significantly reduced using second-order derivatives — particularly if the loss is well approximated by a quadratic — computing the Hessian and its inverse can be computationally demanding as the size of the Hessian scales quadratically with the size of $\boldsymbol{\theta}$. As such, quasi-Newton methods, such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, address some of these limitations by utilizing approximations of the Hessian or the inverse Hessian.

In many cases, the loss function is either non-differentiable or impractical to compute, and derivative-free optimization methods are necessary. One possible method is to use gradient

descent by computing the numerical gradient at each iteration, which can be computed by the central difference method:

$$\nabla_{\boldsymbol{\theta}} l(\boldsymbol{\theta})_i \approx \frac{l(\boldsymbol{\theta} + \epsilon \boldsymbol{e}_i) - l(\boldsymbol{\theta} - \epsilon \boldsymbol{e}_i)}{2\epsilon} \tag{1.12}$$

where we approximate the gradient by computing the loss at a small (ϵ) steps (forward and back) in each coordinate direction of $\boldsymbol{\theta}$. Here, $\boldsymbol{e}_i$ is the elemental vector that contains zeros except for a 1 placed at position i (alternatively, a Kronecker delta function at position i). Therefore, for $\boldsymbol{\theta}$ of size M, this requires 2M loss function evaluations at each iteration.

Another common method for derivative-free optimization is Powell's method. At each iteration, the current estimate is sequentially updated using line-searches performed along a list of M basis vectors (initialized as the elemental vectors). At the end of each iteration, one of the basis vectors is replaced by the direction of total change after the M update steps. Due to the M updates at each iteration — each requiring a line-search — Powell's method can also be very computationally demanding.

Another derivative-free strategy is a stochastic approach called the covariance matrix adaptation-evolution strategy (CMA-ES). In this strategy, a Gaussian sampling pattern is utilized to randomly sample many candidate solutions for $\boldsymbol{\theta}$ at each iteration. The mean vector and covariance matrix for the Gaussian sampling are then updated based on the distribution of the loss function values for these candidate solutions (as well as incorporating the history over previous iterations). Due to the sampling strategy, the method can be robust in the presence of local optima; however, careful consideration must be taken to ensure robust performance, as

the stochastic nature of the method can yield different solutions across multiple runs even if the same initialization is used.

Local optima (i.e., false solutions) are always a concern in image registration, and each of the optimization methods above can be sensitive to them. Strategies to improve robustness include multi-resolution pyramid approaches, where optimization is performed repeatedly in a coarse-to-fine manner with respect to image spatial resolution (achieved by down-sampling and blurring the image). Other strategies involve multi-start approaches, where several optimizations are performed in parallel, each with a different initialization, and the most optimal solution (i.e., that with the strongest loss function) is selected.

1.3.4 Evaluation of Image Registration

Evaluating the performance of image registration is an important aspect of experimental validation and characterization of a registration method. The gold standard for performance evaluation is the accuracy of the geometric transformation parameters themselves, where a mean-squared-error (MSE) analysis of the estimated vs. ground truth transformation parameters is performed. A careful experimental protocol is required for such an analysis, and therefore it is most often performed only in rigid or simulated settings.

Target registration error (TRE) is more commonly used for measuring registration error in real (i.e., non-simulated) data for which the true transformation is unknown. To compute TRE, distinct landmarks are defined in each image, and the distance between corresponding landmarks is computed after registration. Careful consideration must be taken when selecting landmarks to ensure that the landmark locations are unambiguous (with little variability after repeated selection) and pertinent to the interventional task.

Alternatively, registration performance can be evaluated with respect to anatomical segmentations of pertinent anatomy defined within each image. In this case, the registration may be evaluated by comparing the overlap of the segmentations following registration through metrics such as Dice or Hausdorff distances — where Dice quantifies the extent of overlap between two segmentations, and Hausdorff distance quantifies the maximum discrepancy between the segmentation boundaries.

Lastly, another important metric to consider is the failure rate of the registration method. Many registration methods are highly sensitive to false optima, where arbitrarily high registration errors are observed and measurement of geometric accuracy is not particularly meaningful. In this case, robustness of the registration method can be examined by defining a threshold in error for registration failure and examining the failure rate.

1.4 Estimation Theory and Image Registration

In Sec. 1.2.2, figures of merit were presented for evaluation of an image with respect to the task of detection, thus allowing optimization of image acquisition and post-processing parameters with respect to that task. In a similar manner, this dissertation relates such parameters to the task of registration; however, while the detection framework in Sec. 1.2.2 for evaluating image performance with respect to detectability draws from SDT, for image registration we draw from estimation theory to quantitatively understand how factors such as noise and spatial resolution affect the ability to accurately estimate the transformation parameters in image registration.

The work appearing in Section 1.4 was reported in the following conference proceeding and journal papers: (M.D. Ketcha et al., *Proc. SPIE Medical Imaging*, 10135, 2017) [43] and (M.D. Ketcha et al., *IEEE Trans. Med. Imaging*., 36(10), 2017) [44].

1.4.1 Cramer-Rao Lower Bound

One of the fundamental figures of merit in estimation theory is the Cramer-Rao lower bound (CRLB), as it provides a theoretical statistical limit on the expected error for an estimator. In the case that the estimator is unbiased, the CRLB matrix (C_{LB}) is given by the inverse of the Fisher Information Matrix (FIM), which in turn is derived from the log-likelihood function, $\log L(\boldsymbol{I}|\,\boldsymbol{\theta})$, of the data ($\boldsymbol{I}$) conditioned on the parameter vector ($\boldsymbol{\theta}$) being estimated. The FIM examines the curvature of the likelihood function with respect to changes in $\boldsymbol{\theta}$, representing the intuitive concept that if the likelihood function is highly sensitive to perturbations in $\boldsymbol{\theta}$, then $\boldsymbol{\theta}$ can be estimated more accurately. Note that the FIM itself is independent of the estimator and bias. By definition [35], we have:

$$\begin{aligned}[FIM]_{ij} &= E\left\{\frac{\partial \log\,(I|\theta)}{\partial\theta_i}\frac{\partial \log\,(I|\theta)}{\partial\theta_j}\right\} \\ &= -E\left\{\frac{\partial^2 \log L(\boldsymbol{I}|\,\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right\}\end{aligned} \tag{1.13}$$

where expectation (E) is taken over log-likelihood partial derivative terms evaluated at the true solution. The second equality holds under conditions satisfying interchangeable differentiation and integration of the log-likelihood function.

1.4.2 Application to Image Registration

In this section, a simplified 2D translation-only rigid registration scenario is examined to show how the estimation theory framework may be applied to image registration. In the context of Eq. (1.13), $\boldsymbol{I}$ refers to the image data that we are registering, and $\boldsymbol{\theta}$ is the vector of geometric transformation parameters that the registration process attempts to estimate. The two images to be registered are denoted I_1 and I_2, and for ease of derivation we first assume a noiseless I_1 and therefore the signal-known-exactly (SKE) scenario described by the following problem definition:

$$I_1[x, y] = g(x, y) \tag{1.14}$$

$$I_2[x, y] = g(x - u, y - u) + n_2(x, y) \tag{1.15}$$

where g is the true underlying image function, and n_2 is additive white Gaussian noise (AWGN) that is independent of g and has variance σ^2. Note that I_2 is formed with a translation-only displacement of g, with the transformation $\boldsymbol{\theta} = [u, v]^t$ representing the unknown translation between the two images that we are estimating. Furthermore, the $[\cdot]$ matrix notation in $I_i[x, y]$ highlights the process of discretely sampling the continuous underlying image functions.

Given this problem definition, we follow a conceptually similar derivation presented by Kay [35] [which treated 1D time delay estimation (TDE)] to define the log-likelihood function.

We first note that the subtraction of the images at the true shift leaves only the AWGN term; therefore, $L(\boldsymbol{I}|\,\boldsymbol{\theta})$ is simply the product of Gaussian probability density functions:

$$\begin{aligned}\log L(\boldsymbol{I}|\,\boldsymbol{\theta}) &= \log\left(\prod\nolimits_{x,y} c\exp\left(\frac{-\big(I_2(x-u,y-v)-I_1(x,y)\big)^2}{2\sigma^2}\right)\right)\\ &= \sum_{x,y}\frac{-1}{2\sigma^2}[I_2(x-u,y-v)-g(x,y)]^2+const.\end{aligned} \tag{1.16}$$

By noting that the expectation of I_2 evaluated at $\boldsymbol{\theta}$ is g and evaluating the second derivatives, we can compute the FIM from Eq. (1.16) as:

$$\begin{aligned}[FIM]_{ij} &= -E\left\{\frac{\partial^2\log L(\boldsymbol{I}|\,\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right\}\\ &= -E\left\{\frac{\partial}{\partial\theta_i}\left(\frac{-1}{2\sigma^2}\sum_{x,y}2\,[I_2(x-u,y-v)-g(x,y)]\left(\frac{\partial I_2(x-u,y-v)}{\partial\theta_i}\right)\right)\right\}\\ &= -E\left\{\frac{-1}{\sigma^2}\sum_{x,y}[I_2(x-u,y-v)-g(x,y)]\left(\frac{\partial^2 I_2(x-u,y-v)}{\partial\theta_j\partial\theta_i}\right)\right.\\ &\qquad\left.+\left(\frac{\partial I_2(x-u,y-v)}{\partial\theta_j}\right)\left(\frac{\partial I_2(x-u,y-v)}{\partial\theta_i}\right)\right\}\\ &= \frac{1}{\sigma^2}\sum_{x,y}\left(\frac{\partial g(x-u,y-v)}{\partial\theta_i}\right)\left(\frac{\partial g(x-u,y-v)}{\partial\theta_j}\right)\end{aligned} \tag{1.17}$$

where for the case of translation $\boldsymbol{\theta}=[u,v]$, and $\frac{\partial g}{\partial\theta_i}$ can be shown to be the image derivative with respect to the translation direction, $\frac{\partial g(x-u,y-v)}{\partial u}=\left[\frac{-\partial g(m,n)}{\partial m}\right]_{\substack{m=x-u\\ n=y-v}}=-g_x(m,n)$, (and

similarly for v, y), where $g_x(m,n)$ is the partial derivative image with respect to x. Therefore the FIM in this case is:

$$FIM_{SKE,\ AWGN} = \frac{1}{\sigma^2}\begin{bmatrix} \sum_{x,y} {g_x}^2 & \sum_{x,y} g_x\, g_y \\ \sum_{x,y} g_x\, g_y & \sum_{x,y} {g_y}^2 \end{bmatrix} \tag{1.18}$$

Examining the FIM for this scenario, we see that the information associated with a registration task depends generally on two primary components: (1) the image noise (i.e., variance, which is governed largely by image acquisition technique factors such as the level of radiation dose); and (2) the sum-of-squared image gradients (which are governed by the contrast and frequency content of the subject). While this formulation provides useful basic insight, it is limited in that it does not account for the presence of noise in both the I_1 (fixed) and I_2 (moving) images; nor does it account for correlated noise. As shown later in this dissertation, we address this limitation by deriving the lower bound for the scenario in which the noise terms have different magnitude and frequency content.

1.5 Outline and Overview of the Dissertation

1.5.1 Thesis Statement

The thesis of this work is that ***statistical modeling of the factors that govern image registration accuracy provides a foundation for understanding the fundamental limits of registration accuracy and a quantitative guide to selecting imaging protocols and registration parameters that maximize the performance of image registration algorithms***.

The work addresses factors such as the image quality in moving and fixed images, the magnitude of deformation between moving and fixed images, and the statistical diversity of images used in training algorithms based on artificial neural networks. Analogies are drawn to seminal work in SDT and task-based imaging, which conventionally address factors related to tasks of detection or discrimination — related in this work instead to the task of registration. The work involves analysis from first principles of image statistics, testing in a variety of simulations, and validation in physical experiments in the context of image-guided surgery.

1.5.2 Aim 1: Develop a Statistical Model Relating Image Quality to Image Registration Accuracy

In Chapter 2, a statistical model is derived that relates image quality to image registration accuracy. In doing so, two statistical figures of merit for registration performance are realized — analogous to the prewhitening and non-prewhitening observer models for signal detection — which reveal the dependence of registration accuracy on fundamental factors of image quality, namely the MTF and NPS. The model is used to optimize image acquisition protocol (viz., dose) and post-processing parameters (viz., post-processing filters) for the task of image registration. The work in Chapter 2 was presented at the following conference:

> M.D. Ketcha, T. De Silva, R. Han, A. Uneri, J. Goerres, M.W. Jacobson, S. Vogt, G. Kleinszig, and J.H. Siewerdsen, "Fundamental limits of image registration performance: effects of image noise and resolution in CT-guided interventions," *SPIE Medical Imaging*, Orlando, FL, Oral Presentation (February 2017).

and published in the following journal and conference proceedings:

> M. D. Ketcha, T. De Silva, R. Han, A. Uneri, J. Goerres, M. W. Jacobson, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, "Effects of Image Quality on the Fundamental Limits

of Image Registration Accuracy," *IEEE Trans. Med. Imaging*, vol. 36, no. 10, pp. 1997–2009, 2017.

M. Ketcha, T. De Silva, R. Han, A. Uneri, J. Goerres, G. G, S. Vogt, and J. Siewerdsen, "Fundamental limits of image registration performance: Effects of image noise and resolution in CT-guided interventions," in *Proc. SPIE*, 2017, vol. 10135, p. 1013508.

1.5.3 Aim 2: Develop a Statistical Model for Soft-Tissue Deformation in Rigid Image Registration

In Chapters 3–4, the framework of Aim 1 is extended to model the confounding influence of soft-tissue deformation on rigid image registration. In a manner analogous to background clutter in SDT, soft-tissue deformation is shown to behave as a noise source in rigid image registration and can be modeled according to a power-law distribution. From this model, it is shown theoretically and experimentally that gradient-based similarity metrics are relatively robust to the effect of deformation compared to histogram-based similarity measures. Furthermore, Chapter 4 demonstrates how insights from this model may be applied to 3D-2D (CT-to-radiograph) registration in spine surgery, where global spine deformation must be accounted in a locally rigid manner. The work in Chapters 3–4 was presented at the following conferences:

M.D. Ketcha, T. De Silva, A. Uneri, G. Kleinszig, S. Vogt, J.P. Wolinsky, and J.H. Siewerdsen, "Automatic masking for robust 3D-2D image registration in image-guided spine surgery," *SPIE Medical Imaging*, San Diego, CA, Oral Presentation (February 2016).

M.D. Ketcha, T. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J.H. Siewerdsen, "A statistical model for image registration performance: effect of tissue deformation." *SPIE Medical Imaging*, Houston, TX, Oral Presentation (February 2018).

M.D. Ketcha, T. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J.H. Siewerdsen, "A Statistical Model Relating Image Quality to Image Registration Accuracy in Image-Guided Surgery," *Bull. Am. Phys. Soc.*, Boston, MA Oral Presentation (March 2019).

and published in the following journal and conference proceedings:

M. D. Ketcha, T. De Silva, A. Uneri, M. W. Jacobson, J. Goerres, G. Kleinszig, S. Vogt, J. P. Wolinsky, and J. H. Siewerdsen, "Multi-stage 3D-2D registration for correction of anatomical deformation in image-guided spine surgery," *Phys. Med. Biol.*, vol. 62, no. 11, p. 4604, 2017.

M. D. Ketcha, T. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, "A Statistical Model for Rigid Image Registration Performance : The Influence of Soft-Tissue Deformation as a Confounding Noise Source," *IEEE Trans. Med. Imaging*, vol. 38. no. 9, pp. 2016-2027, 2019.

1.5.4 Aim 3: Investigate the Effect of Statistical Mismatch for CNN-Based Methods

In Chapter 5, factors of image spatial resolution and noise are considered in relation to registration algorithms based on a convolutional neural network (CNN). Specifically, the work investigates how the statistical properties of images in the training set affects the ability of the network to generalize to test data with differing statistical characteristics. The analysis yields insight on the benefits of statistically diverse training data and is applied to scenarios of both CNN-based deformable image registration and CNN-based MR-to-CT synthesis. The work in Chapter 5 was presented at the following conferences:

M. D. Ketcha, T. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J.H. Siewerdsen, "Effect of statistical mismatch between training and test images for CNN-based deformable registration." *SPIE Medical Imaging*, San Diego, CA, Oral Presentation (February 2019)

M.D. Ketcha, C.K. Jones, P. Wu, R. Han, A. Uneri, J. Lee, M. Luciano, W.S. Anderson, and J.H. Siewerdsen. "Deformable MR to Cone-Beam CT Registration for High-

Precision Neuro-Endoscopic Surgery." National Image Guided Therapy Workshop 2020. Virtual Conference, Poster Presentation (April 2020).

and published in the following journal and conference proceedings:

M. D. Ketcha, T. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, “Effect of statistical mismatch between training and test images for CNN-based deformable registration,” in *Proc. SPIE*, 2019, vol. 10949, p. 109490T.

M. D. Ketcha, T. S. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, “Learning-based deformable image registration: effect of statistical mismatch between train and test images,” *J. Med. Imaging*, vol. 6, no. 4, p. 44008, 2019.

In Chapter 6, the key findings of this work are reviewed and interpreted within related theoretical contexts, and the implications for ongoing research on image registration methods for image-guided interventions are discussed. Assumptions and limitations of the work are acknowledged, and potential future work is outlined.

Chapter 2: Effects of Image Quality on the Fundamental Limits of Image Registration Accuracy

2.1 Introduction

In image-guided interventions, registration performance pertains to the accuracy with which the preoperative image (denoted I_1) and intraoperative image (denoted I_2) can be co-registered in a common coordinate system. In many scenarios, the ability to accurately register I_1 to I_2 (and planning data therein) may be even more important than the ability to visualize structures in I_2 directly. In image-guided neurosurgery, for example, an intraoperative CT/CBCT image may have image quality insufficient for direct visualization of the surgical target, but it provides sufficient structural information to drive registration of a preoperative CT or MRI in which the target has been clearly visualized and delineated. In this context, the primary task relates to registration more so than direct visualization, and it is important to understand how the image quality of the intraoperative image affects the accuracy of registration.

A further example can be found in Uneri *et al*. [36], who developed a registration method to evaluate surgical screw placement relative to preoperative CT, enabling quantitative evaluation of screw malplacement in 3D rather than qualitative visualization and interpretation of 2D projection radiographs. In this scenario as well, the primary task is accurate registration to CT (and overall perceptual image quality in the radiographs is of secondary importance). In related work, Uneri *et al*. [37] reported that accurate registration could be achieved even when the radiograph was acquired at a dose ~1/10th that of standard technique (a dose at which quantum noise largely prohibits clear visual interpretation), indicating that the task of registration may be more robust against noise than the task of visualization; hence, imaging parameters that are optimal for visualization and detectability may not correspond to those that are optimal for registration.

The development of imaging systems for interventional guidance therefore prompts consideration that the optimal I_2 imaging technique (i.e., factors governing image noise and spatial resolution in the intraoperative image) is that which provides a desired level of registration accuracy, rather than visual image quality. This consideration in turn motivates a quantitative framework to relate registration accuracy to image quality.

As discussed in Section 1.2, image quality metrics such as MTF, NPS, and NEQ (which are particularly well quantified in CT and CBCT imaging) not only provide a rich characterization of spatial resolution and noise but also have been shown to reliably guide image acquisition and reconstruction techniques with respect to tasks of detection and visualization [12], [16], [20], [38]–[42]. However, while CT and CBCT images are increasingly employed in image-guided interventions, there is comparatively little rigorous understanding of the relationship between these image quality factors and registration

performance, leaving largely unanswered fundamental questions in determining imaging techniques that achieve a desired level of registration accuracy.

Registration performance is commonly investigated by rigorous measurement and experimental evaluation of geometric accuracy in contexts appropriate to a particular application. Such investigation often involves registration repeated for either a large data set or simulated noise realizations, where the output transformation parameters are compared to the ground truth transformation. Results from such experiments provide an important basis for quantifying performance in support of the clinical application; however, they are still often performed under the general assumption that a "higher quality" image will give better registration performance — or that a level of image quality sufficient for visualization will in turn be sufficient for registration — without rigorous guidance of an analytical model to support such an assumption. As a result, there are untested opportunities for imaging methods that are best suited for the task of registration — e.g., methods that achieve a desired level of registration accuracy with reduced radiation dose.

In this chapter, we seek an analytical framework that will help to unify models of image quality (e.g., spatial resolution and noise) with models for registration performance, providing a rigorous basis and guide to selection of image acquisition protocols, reconstruction methods, and post-processing techniques sufficient (or optimal) for the task of registration. We approach the question by analyzing a simple model involving 2D translation-only registration to gain initial insight into the more complicated general registration problem. As detailed below, we build from well-established, image quality considerations for CT/CBCT image quality [6]–[10] and realize a framework that relates these factors to the task of image registration. While this framework is general to the broader field of image registration, it is presented here

specifically in the context of medical imaging, relating registration performance to concepts of image noise, spatial resolution, and information-theoretic metrology (viz., image NPS, MTF, and NEQ) that are familiar and prevalent in medical image quality assessment — particularly in x-ray CT and CBCT.

The work appearing in this chapter was reported in the following conference proceeding and journal papers: (M.D. Ketcha et al., *Proc. SPIE Medical Imaging,* 10135, 2017) [43] and (M.D. Ketcha et al., *IEEE Trans. Med. Imaging.,* 36(10), 2017) [44].

2.2 Statistical Evaluation of Image Registration

2.2.1 Prior Work

Over the past 15 years, there have been several contributions to understanding the lower bounds in image registration accuracy. Robinson and Milanfar performed early work in statistical evaluation of registration performance by deriving the CRLB for translation-only image registration in the presence of AGWN. Yetik and Nehorai [45] extended this derivation to model both translation and rotation, and Pham *et al.* [46] extended the model to more general projective transformations. Uss *et al.* [47], through an alternative approach, derived a translation-only CRLB that assumed AGWN and an underlying image distributed according to a fractal Brownian motion model, showing good agreement with measurements of registration performance in high SNR scenarios. Xu *et al.* [48] derived Ziv-Zikai Bounds (ZZB) for translation-only registration and was able to model the steep drop in performance in very low SNR conditions due to registration failure. Aguerrebere *et al.* [49] later explained these works to be associated with the SKE and derived the CRLB for registering two or more

images, each of which contained stationary Gaussian noise (no longer limited to AGWN). Their work also examined various other lower bounds such as the extended ZZB for white noise contaminated images and a Bayesian CRLB when a shift prior was known.

Beyond evaluation of the lower bounds, it is also important to examine the registration method itself, which includes factors such as image preprocessing, similarity metric, and optimization method. Aguerrebere et al. [50] provided a review of general registration frameworks (particularly in the presence of white noise) and distinguished methods that do not rely on prior information (e.g., Maximum Likelihood Estimator) from those that do by incorporating information about the statistical distribution of both the signal and noise (e.g., Bayesian Maximum Likelihood Estimator via the Wiener filter). Robinson and Milanfar [51] and Pham *et al.* [46] demonstrated the bias present in several registration estimators, fundamentally limiting the registration performance in very high SNR scenarios. The effect of image quality on registration accuracy was investigated by Zhao et al. [52] for translation-only registration under AGWN to understand the influence of spatial resolution (cf., noise) on the SSD similarity metric. Their work indicated that when registering images of different spatial resolution using SSD, the higher resolution image should be blurred to match the lower resolution. The result is particularly interesting, since by the data-processing inequality, such blur does not improve the CRLB and thus depends on the similarity metric itself (whereas the CRLB is independent of the similarity metric). Such work demonstrates the necessity for a statistical registration framework to examine both the theoretical limits of registration accuracy and the influence of the similarity metric to more fully describe the relationship between image quality and registration performance.

2.2.2 CRLB for Image Guided Interventions

2.2.2.1 Derivation of the CRLB

A typical scenario in image-guided procedures involves the registration of a high-quality preoperative image to a lower-quality (noisy and/or blurry) intraoperative image, requiring a formulation that explicitly addresses the presence of noise in both images and allowing the noise in each image to carry disparate (heteroscedastic) magnitude and correlation. For this we take the following image model:

$$I_1[x,y] = g(x,y) + n_1(x,y) \tag{2.1}$$

$$I_2[x,y] = g(x-u, y-v) + n_2(x,y) \tag{2.2}$$

where we take a similar formulation as Equations (1.14) and (1.15) in that we are estimating the transformation $\boldsymbol{\theta} = [u,v]^t$ among these discretely sampled images (denoted with square bracket matrix notation, $I_i[x,y]$); however, we now have noise present in both images, and neither is constrained to AWGN. We assume linear systems with stationary Gaussian signal and noise, where n_i is zero-mean and independent of g and $n_{j\neq i}$. Such assumptions are a common starting point in formulation of image statistics that can be extended to 'local' approximation for nonlinear, nonstationary systems (discussed below and in [40], [53]). Below, we derive the CRLB for this scenario, analogous to that derived for 1D TDE [54]–[56].

Due to the presence of correlated noise terms in both images, we lose the simple form for the likelihood function presented in Sec 1.4.2, causing the spatial domain analysis to become less tractable. A Fourier representation, however, will facilitate the analysis. We take $Z_1[f_x, f_y]$ and $Z_2[f_x, f_y]$ as the 2D Fourier transforms ($\mathcal{F}\{\cdot\}$) of I_1 and I_2, respectively, where

for purposes of the proof, we break from the convention of the $(\cdot)$ notation for continuous Fourier domain functions to directly represent the discrete Fourier transform of the image. We first note that for a signal bandlimited below f_{Nyq} (i.e., no aliasing), Z_1 comprises the sampled frequency representation of the signal and noise so that:

$$Z_1[f_x, f_y] = \mathcal{F}\{g(x,y)\}(f_x, f_y) + \mathcal{F}\{n_1(x,y)\}(f_x, f_y) \tag{2.3}$$

and by the shift property we have:

$$Z_2[f_x, f_y] = e^{-j2\pi(f_x u + f_y v)}\mathcal{F}\{g(x,y)\}(f_x, f_y) + \mathcal{F}\{n_2(x,y)\}(f_x, f_y) \tag{2.4}$$

If we examine a particular frequency location $[f_x^{(m)}, f_y^{(n)}]$ (where m, n refer to the indexed frequency samples) we see that the covariance of $Z_1\left[f_x^{(m)}, f_y^{(n)}\right], Z_2\left[f_x^{(m)}, f_y^{(n)}\right]$ is:

$$\begin{aligned} K_{mn} &= E\{X_{mn}X_{mn}^* | \boldsymbol{\theta}\} \text{ where } X_{mn} = \begin{bmatrix} Z_1\left[f_x^{(m)}, f_y^{(n)}\right] \\ Z_2\left[f_x^{(m)}, f_y^{(n)}\right] \end{bmatrix} \\ &= \begin{bmatrix} G\left[f_x^{(m)}, f_y^{(n)}\right] + N_1\left[f_x^{(m)}, f_y^{(n)}\right] & G\left[f_x^{(m)}, f_y^{(n)}\right] e^{-j2\pi(f_x^{(m)}u + f_y^{(n)}v)} \\ G\left[f_x^{(m)}, f_y^{(n)}\right] e^{j2\pi(f_x^{(m)}u + f_y^{(n)}v)} & G\left[f_x^{(m)}, f_y^{(n)}\right] + N_2\left[f_x^{(m)}, f_y^{(n)}\right] \end{bmatrix} \end{aligned} \tag{2.5}$$

where G and N_i are the power spectra of g and n_i, respectively, and * denotes the complex conjugate transpose. With this in mind, we wish to represent the entirety of the data in a single vector. Therefore, we concatenate Z_1 and Z_2 into the vector $\boldsymbol{X}$ in a manner that the corresponding frequencies in Z_1 and Z_2 are adjacent to each other:

$$\begin{aligned} \boldsymbol{X} = [\ & Z_1\left[f_x^{(1)}, f_y^{(1)}\right], Z_2\left[f_x^{(1)}, f_y^{(1)}\right], \ldots Z_1\left[f_x^{(M)}, f_y^{(1)}\right], Z_2\left[f_x^{(M)}, f_y^{(1)}\right], \\ & Z_1\left[f_x^{(1)}, f_y^{(2)}\right], Z_2\left[f_x^{(1)}, f_y^{(2)}\right], \ldots Z_1\left[f_x^{(M)}, f_y^{(N)}\right], Z_2\left[f_x^{(M)}, f_y^{(N)}\right]\ ]^t \end{aligned} \tag{2.6}$$

where $f_x^{(i)}, f_y^{(i)}$ are the indexed frequency samples $\in [-f_{Nyq},\ f_{Nyq}]$. Under the assumption

of stationary signal and noise, the frequency components of Z_i are statistically independent [57], and the covariance matrix of $\boldsymbol{X}$ has a block diagonal form:

$$K = E\{\boldsymbol{X}\boldsymbol{X}^* | \boldsymbol{\theta}\} = \begin{bmatrix} K_{11} & & & & & \\ & \ddots & & & \mathbf{0} & \\ & & K_{M1} & & & \\ & & & K_{12} & & \\ & \mathbf{0} & & & \ddots & \\ & & & & & K_{MN} \end{bmatrix} \tag{2.7}$$

where the block-diagonal components are described in Eq. (2.6). By assuming both the image function and noise to be zero-mean Gaussian processes, the frequency components of the Fourier representation are also jointly Gaussian, and the likelihood function for $\boldsymbol{X}$ can be written:

$$L(\boldsymbol{X}|\,\boldsymbol{\theta}) = \frac{1}{\det(\pi K)} \exp\left(-\boldsymbol{X}^* K^{-1} \boldsymbol{X}\right) \tag{2.8}$$

As shown in [54], the FIM for this complex Gaussian likelihood function can be reduced to:

$$[FIM]_{ij} = \mathrm{Tr}\left[\left(K^{-1}\frac{\partial K}{\partial \theta_i}\right)\left(K^{-1}\frac{\partial K}{\partial \theta_j}\right)\right] \tag{2.9}$$

Combining Eqs. (2.5), (2.7), and (2.9) gives the 2×2 FIM:

$$\begin{aligned} [FIM]_{ij} &= \frac{1}{2}\sum_{m,n} \mathrm{Tr}\left[\left(K_{mn}^{-1}\frac{\partial K_{mn}}{\partial \theta_i}\right)\left(K_{mn}^{-1}\frac{\partial K_{mn}}{\partial \theta_j}\right)\right] \\ &= \frac{1}{2}\sum_{m,n} \frac{2(2\pi)^2 f_i^{(m)} f_j^{(n)} G^2\left[f_x^{(m)}, f_y^{(n)}\right]}{G\left[f_x^{(m)}, f_y^{(n)}\right] N_1\left[f_x^{(m)}, f_y^{(n)}\right] + G\left[f_x^{(m)}, f_y^{(n)}\right] N_2\left[f_x^{(m)}, f_y^{(n)}\right] + N_2\left[f_x^{(m)}, f_y^{(n)}\right] N_1\left[f_x^{(m)}, f_y^{(n)}\right]} \end{aligned} \tag{2.10}$$

where we have denoted $f_1 = f_x$ and $f_2 = f_y$ in reference to $f_i^{(m)}$ and $f_j^{(n)}$ when computing $[FIM]_{ij}$, and the 1/2 term prior to the sum is included to compensate for symmetry in the Fourier domain (thus over-representing the information by a factor of 2). Given sufficiently

high sampling density, we may approximate the summation as an integral to write Eq. (2.10) as:

$$FIM = (2\pi)^2 A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix}, \tag{2.11a}$$

$$\text{where} \quad \gamma_{ij} = \iint_{-f_{Nyq}}^{f_{Nyq}} \frac{f_i f_j G^2(f_x, f_y)\, df_x df_y}{G(f_x, f_y)N_1(f_x, f_y) + G(f_x, f_y)N_2(f_x, f_y) + N_1(f_x, f_y)N_2(f_x, f_y)} \tag{2.11b}$$

and A is the image area [i.e., the extent of the image in x and y directions]. To simplify notation, we remove explicit (f_x, f_y) dependence:

$$FIM = (2\pi)^2 A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix}, \tag{2.12a}$$

$$\text{where} \quad \gamma_{ij} = \iint_{-f_{Nyq}}^{f_{Nyq}} \frac{f_i f_j G^2}{GN_1 + GN_2 + N_1 N_2} df_x df_y \tag{2.12b}$$

Extension to a pair of 3D images is straightforward, giving:

$$FIM = (2\pi)^2 V \begin{bmatrix} \gamma_{xx} & \gamma_{xy} & \gamma_{xz} \\ \gamma_{xy} & \gamma_{yy} & \gamma_{yz} \\ \gamma_{xz} & \gamma_{yz} & \gamma_{zz} \end{bmatrix}, \tag{2.13a}$$

$$\text{where} \quad \gamma_{ij} = \iiint_{-f_{Nyq}}^{f_{Nyq}} \frac{f_i f_j G^2}{GN_1 + GN_2 + N_1 N_2} df_x df_y df_z \tag{2.13b}$$

where V is the image volume.

2.2.2.2 Examination of the CRLB

Equation (2.12) provides the necessary framework to analyze registration performance bounds when the two images have separate noise forms that are not necessarily AGWN. We

rewrite Eq. (2.12b) as follows to more explicitly show the dependence of Fisher information on the image signal and noise:

$$FIM = (2\pi)^2 A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix}, \tag{2.14a}$$

$$\text{where} \quad \gamma_{ij} = \iint_{-f_{Nyq}}^{f_{Nyq}} f_i \, f_j \, SNR \, df_x df_y \tag{2.14b}$$

$$\text{and} \quad SNR(f_x, f_y) = \frac{G^2}{GN_1 + GN_2 + N_1 N_2} \tag{2.15}$$

Equation (2.14) carries intuitive dependencies of registration performance on SNR, namely: the numerator scales with signal feature strength (i.e., contrast and gradient magnitude), and the denominator scales with image noise, including cross terms corresponding to noise in cross-correlation (examined further in Section 2.2.3).

One can see that the general form in Eq. (2.12) reduces to Eq. (1.18) in the simplified case of noiseless I_1 ($N_1 = 0$) and AWGN (N_2 is "white" with magnitude σ^2). From Eq. (2.12) this suggests:

$$FIM_{SKE,\ AWGN} = (2\pi)^2 A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix}, \tag{2.16a}$$

$$\text{where} \quad \gamma_{ij} = \iint_{-f_{Nyq}}^{f_{Nyq}} f_i f_j \frac{G}{N_2} df_x df_y \tag{2.16b}$$

$$= \frac{1}{\sigma^2} \iint_{-f_{Nyq}}^{f_{Nyq}} f_i f_j \, G \, df_x df_y \tag{2.17}$$

which is equivalent to Eq. (1.18) via Parseval's Theorem and the Fourier derivative theorem.

A simplifying approximation of the FIM is obtained when the denominator of the SNR is not a function of G, as in the simple case of Eq. (2.16b). This form is achieved when $N_1 N_2 \ll GN_1 + GN_2$ allowing approximation of the denominator in Eq. (2.15) as $GN_1 + GN_2$ and giving:

$$FIM_{\widehat{N} \ll G} = (2\pi)^2 A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix}, \tag{2.18a}$$

$$\text{where} \quad \gamma_{ij} = \iint_{-f_{Nyq}}^{f_{Nyq}} f_i \, f_j \, SNR_{\widehat{N} \ll G} \, df_x df_y \tag{2.18b}$$

$$\text{with} \quad SNR_{\widehat{N} \ll G}(f_x, f_y) = \frac{G}{N_1 + N_2} \tag{2.19}$$

The approximation can also be written as $N_1 N_2/(N_1 + N_2) = \widehat{N} \ll G$, showing that this approximation holds when the signal is high compared to the noise. An interesting observation is that this approximation is analytically equivalent to treating the noise as being contained in only one image:

$$I_1[x, y] = g(x, y) \tag{2.20}$$

$$I_2[x, y] = g(x - u, y - u) + n_2(x, y) + n_1(x, y) \tag{2.21}$$

indicating that (for low noise levels) it is not important how the noise is distributed between the two images, and only the sum $(N_1 + N_2)$ affects registration performance. For the simple

AWGN case, Eq. (2.18) amounts to the SKE scenario in Eqs. (1.14) and (1.15) and the FIM becomes:

$$FIM_{\widehat{N}\ll G,\ AWGN} = \frac{1}{{\sigma_1}^2 + {\sigma_2}^2}\begin{bmatrix} \sum_{x,y} {g_x}^2 & \sum_{x,y} g_x g_y \\ \sum_{x,y} g_x g_y & \sum_{x,y} {g_y}^2 \end{bmatrix} \tag{2.22}$$

where we now have the sum of the variances from the two images entering the denominator of the $FIM_{\widehat{N}\ll G,\ AWGN}$. In the case of equal variance (homoscedastic), this amounts to an increase in the CRLB by a factor of 2 compared to Eq. (1.18).

In this reduced form, the $FIM_{\widehat{N}\ll G}$ captures intuitive dependencies between noise and registration performance (i.e., the CRLB scales directly with total variance). However, we will see that it further yields simple relationships for NEQ and DQE (shown in Sec. 2.2.4). The $N_1 N_2/(N_1 + N_2) = \widehat{N} \ll G$ condition corresponds to at least three relevant scenarios: (i) the signal is high compared to the total noise (i.e, $G \gg (N_1 + N_2) > N_1 N_2/(N_1 + N_2)$, indicating that $SNR_{\widehat{N}\ll G}$ is large; (ii) I_1 is noiseless ($N_1 \approx 0$); or (iii) I_2 is much noisier than I_1 (i.e., $N_1 \ll N_2$) *and* the signal power is high compared to the noise in I_1 (i.e., $N_1 \ll G$). These are suitable approximations, for example, in registration of high-contrast bone anatomy (high signal power in g) from a high quality preoperative CT (low N_1) to a lower quality (high N_2) intraoperative CBCT.

2.2.3 Maximizing Cross-correlation And Optimal filtering

2.2.3.1 Derivation of Expected Registration Error Using Cross-Correlation

An important question that an analytical registration model must also address is the impact of (optional) post-processing blur. It can be seen in Eq. (2.12b) that blur by a simple linear filter (described by the MTF^2 implicit in both the G and N_i terms) will exactly cancel out for invertible filters (with non-zero MTF), exemplifying the information-theoretic data processing inequality: application of a linear filter does not affect the CRLB in registration performance. However, in practice the benefits of post-processing blur are well known (and shown in Sec. 2.4.3) for reducing the impact of high-frequency noise on registration performance. Therefore, to more fully examine the question of spatial resolution and optimal filtering, the registration method itself must be examined. In this section we focus on the commonly used registration method of maximizing cross-correlation:

$$\begin{aligned} r(\tau,\varphi) &= (I_1 \otimes I_2)(\tau,\varphi) \\ &= \sum_{x,y} I_1[x,y] I_2[x+\tau, y+\varphi] \end{aligned} \tag{2.23}$$

where an interpolation over I_2 must take place to achieve the continuous cross-correlation function. By examining a local region near the peak of the cross-correlation function, we observe that an equivalent estimator of $\boldsymbol{\theta} = [u, v]$ is one that solves for $\widehat{\boldsymbol{\theta}} = [\hat{u}, \hat{v}]$ such that $r_\tau(\hat{u}, \hat{v}) = \frac{\partial r(\hat{u},\hat{v})}{\partial \tau} = 0$ and $r_\varphi(\hat{u}, \hat{v}) = \frac{\partial r(\hat{u},\hat{v})}{\partial \varphi} = 0$. If we assume errors in the estimation to be

contained within the linear region near the true solution (similar to [58] in 1D TDE), a first-order Taylor series approximation near $\boldsymbol{\theta}$ yields the error estimate:

$$\begin{bmatrix}(\hat{u}-u)\\(\hat{v}-v)\end{bmatrix}^t \approx \begin{bmatrix}r_\tau(u,v)\\r_\varphi(u,v)\end{bmatrix}^t \begin{bmatrix}r_{\tau\tau}(u,v) & r_{\tau\varphi}(u,v)\\r_{\tau\varphi}(u,v) & r_{\varphi\varphi}(u,v)\end{bmatrix}^{-1} \tag{2.25}$$

where $r_{\tau\tau}(u,v) = \frac{\partial^2 r(u,v)}{\partial \tau^2}$, and similarly for the other 2nd derivative terms. The root-mean-squared error (RMSE) of the estimate is obtained by computing the magnitude of the expectation of this error.

We begin by simplifying Eq. (2.25) by assuming $r_{\tau\varphi}(u,v)$ to be small in comparison to the diagonal terms, giving:

$$(\hat{u}-u) \approx r_\tau(u,v)/r_{\tau\tau}(u,v) \tag{2.26}$$

$$(\hat{v}-v) \approx r_\varphi(u,v)/r_{\varphi\varphi}(u,v) \tag{2.27}$$

By the associative property of cross correlation:

$$\begin{aligned} r(u,v) &= (I_1 \otimes I_2)(u,v) \\ &= (g\otimes g)(0,0) + (g\otimes n_2)(u,v) + (n_1\otimes g)(0,0) + (n_1\otimes n_2)(u,v) \end{aligned}$$

and without loss of generality:

$$r(u,v) = (g\otimes g)(0,0) + (g\otimes n_2)(0,0) + (n_1\otimes g)(0,0) + (n_1\otimes n_2)(0,0) \tag{2.28}$$

where the second equality carries the discretization of g and n_i in forming I_i as well as the implicit $[u,v]$ shift of g in I_2 so that $g\otimes g$ and $n_1\otimes g$ are evaluated at $(0,0)$. Further, this equality neglects interpolation and assumes an un-aliased signal. The third equality, which does not affect the result (as the expectation is unaffected), is introduced for notational convenience. From examination of Eq. (2.28), we note that the numerators of Eqs. (2.26) and (2.27) have expected value equal to zero. Furthermore, by definition $\partial(g\otimes g)/\partial\theta_i\,(0,0) = 0$.

Therefore, near solution the numerators will comprise the remaining terms in the derivative of Eq. (2.28). On the other hand, the expectation of the denominator is non-zero and depends only on the signal term, indicating that near solution the denominator is primarily determined by the signal term alone. Therefore, we have Eq. (2.26) to first-order approximation:

$$\begin{aligned}(\hat{u}-u) &\approx \frac{\frac{\partial}{\partial\tau}(g\otimes n_2 + n_1\otimes g + n_1\otimes n_2)(0,0)}{\frac{\partial^2(g\otimes g)}{\partial\tau^2}(0,0)} \\ &= \frac{\sum_{m.n}\left(j2\pi f_x^{(m)}\right)(F_g\bar{F}_{n_2} + F_{n_1}\bar{F}_g + F_{n_1}\bar{F}_{n_2})}{\sum_{m,n}\left(j2\pi f_x^{(m)}\right)^2 F_g\bar{F}_g}\end{aligned} \tag{2.29}$$

with the second equality following from the Fourier derivative property and Parseval's Theorem, where $F_g\left[f_x^{(m)}, f_y^{(n)}\right]$ denotes the Fourier coefficients of the discretized g (similarly for F_{n_i}), where ${f_x}^{(m)}, {f_y}^{(n)}$ are the indexed frequency samples $\in [-f_{Nyq},\ f_{Nyq}]$, and the hat denotes complex conjugation. Explicit notation of the frequency dependence on the F_i terms is excluded for notational convenience in Eq. (2.29) and the equations below. As the expectation of the error is zero, we need only to examine the variance:

$$\mathrm{Var}(\hat{u}) \approx \frac{\mathrm{Var}\left(\sum_{m.n}\left(jf_x^{(m)}\right)(F_g\bar{F}_{n_2} + F_{n_1}\bar{F}_g + F_{n_1}\bar{F}_{n_2})\right)}{(2\pi)^2\left[\sum_{m,n}\left(f_x^{(m)}\right)^2 F_g\bar{F}_g\right]^2} \tag{2.30}$$

From the assumption of stationarity, the frequency components of the Fourier terms in the numerator are independent [57], leaving only the sum over the variance terms:

$$\mathrm{Var}(\hat{u}) \approx \frac{\sum_{m,n}\left(f_x^{(m)}\right)^2 E\{(F_g\bar{F}_{n_2} + F_{n_1}\bar{F}_g + F_{n_1}\bar{F}_{n_2})(\bar{F}_g F_{n_2} + \bar{F}_{n_1}F_g + \bar{F}_{n_1}F_{n_2})\}}{(2\pi)^2\left[\sum_{m,n}\left(f_x^{(m)}\right)^2 F_g\bar{F}_g\right]^2} \tag{2.31}$$

where $G\left[f_x^{(m)}, f_y^{(n)}\right]$ and $N_i\left[f_x^{(m)}, f_y^{(n)}\right]$ are the signal and noise-power spectra, respectively. By analogous derivation for $\mathrm{Var}(\hat{v})$, and approximating the sum as an integral, the RMSE is:

$$RMSE \approx \sqrt{1/\rho_x + 1/\rho_y}\text{ , where}$$

$$\rho_i = \frac{(2\pi)^2 A\left[\iint_{-f_{Nyq}}^{f_{Nyq}} f_i^2 G\, df_x df_y\right]^2}{\iint_{-f_{Nyq}}^{f_{Nyq}} f_i^2 (GN_1 + GN_2 + N_1N_2)\, df_x df_y} \tag{2.32}$$

and A is the image area as described above.

2.2.3.2 Examination of the RMSE Estimate

The result for RMSE in Eq. (2.32) carries many similarities to the FIM in Eq. (2.12), particularly with respect to the noise term $GN_1 + GN_2 + N_1N_2$. From the associative property of cross correlation, we see that $I_1 \otimes I_2 = g\otimes g + g\otimes n_2 + n_1\otimes g + n_1\otimes n_2$, which comprises two primary terms: (i) the true cross correlation of the signal $g\otimes g$; and (ii) the remaining terms associated with additive noise in the cross-correlation. The power associated with these noise terms is represented in the Fourier domain as $GN_1 + GN_2 + N_1N_2$, which appears in both Eqs. (2.12) and (2.32). Equation (2.32) is particularly interesting in that the signal and noise terms (numerator and denominator of ρ_i, respectively) are integrated separately, analogous to a non-

prewhitening detectability model (discussed in Sec. 1.2.2), for which an "optimal blur" can be derived — distinct from the simple data processing concept noted above. Noting that blurring the images leads to convolution in the cross-correlation [i.e., $(h_1 * I_1)\otimes(h_2 * I_2) = (h_1\otimes h_2) * (I_1\otimes I_2) = h * r$], we may consider a blur kernel (h) with Fourier transform H, giving:

$$RMSE \approx \sqrt{1/\rho_x + 1/\rho_y}\ , \text{ where}$$

$$\rho_i = \frac{(2\pi)^2 A \left[\iint_{-f_{Nyq}}^{f_{Nyq}} f_i^2 HG\ df_x df_y\right]^2}{\iint_{-f_{Nyq}}^{f_{Nyq}} f_i^2 H^2 (GN_1 + GN_2 + N_1N_2)\ df_x df_y} \tag{2.33}$$

This result then can be minimized as a function of H. In the 1D TDE case, Knapp *et al.* [58] showed that the maximum likelihood estimate is achieved by filtering according to the data-dependent Hannan and Thompson method, giving:

$$H_{HT} = \frac{G/(N_1N_2)}{1 + G/N_1 + G/N_2} \tag{2.34}$$

In practice, however, blurring of I_1 or I_2 is typically achieved using a simple kernel (e.g., symmetric Gaussian blur); thus, we consider a blurring function H_{CC}:

$$H_{CC}(f_x, f_y) = e^{-4\pi^2(f_x^2 + f_y^2)\sigma_b^2} \tag{2.35}$$

where H_{CC} represents the effect of blurring both images by a Gaussian kernel of width σ_b. Therefore, we can solve for the optimal blur by minimizing Eq. (2.33) with respect to σ_b. Depending on the registration method, additional blur can be included explicitly (such as the blur incurred with interpolation, described by H_{interp}), giving a combined $H = H_{interp}H_{CC}$ for the total blur appearing in Eq. (2.33).

2.2.4 Connections to Image Quality

As derived above, the CRLB for image registration depends explicitly on the noise and spatial resolution characteristics of the imaging system(s). Recent decades have seen the development of accurate models for the image quality characteristics of CT and CBCT imaging systems [6]–[10], including the MTF and NPS, and their dependence on each factor in the imaging chain, such as dose, system geometry, acquisition technique, and reconstruction technique. Such models consider the propagation of signal and noise through the imaging chain to describe the MTF and NPS. Simplifying from the 3D case in [6] to the 2D case considered here, a simple form for the *ideal* axial CT image NPS (i.e., a deterministic system featuring quantum noise and blur, but without aliasing or electronic noise) can be written:

$$NPS_{ideal}(f_x, f_y) = \frac{\pi\sqrt{f_x^2 + f_y^2}}{m\bar{q}M^2\Gamma} MTF^2(f_x, f_y) \tag{2.36}$$

(with the full form of the NPS detailed in [6]) where the dose is related to the number of projections (m) and incident x-ray fluence ($\bar{q}$), M refers to system magnification, and Γ is the system gain. Considering the SKE white noise registration model of Eq. (1.18) and the image quality model of Eq. (2.36), we reach an immediate finding: since noise power [and thus total noise variance, given by the integral of Eq. (2.36)] is inversely proportional to dose (via the $m\bar{q}$ term in the denominator), and the CRLB is proportional to noise variance, then the lower bound on registration error scales in inverse proportion to dose. Incorporating noise terms in both images shows a more complex relationship that depends not only on the total dose, but the relative dose in each image.

We can consider such relationships further in terms of metrics of fidelity that incorporate both the MTF and NPS. The performance of an imaging system is commonly described in terms of the NEQ, representing the effective number of incident photons contributing to each spatial frequency [6]:

$$NEQ(f_x, f_y) = \pi\sqrt{f_x^2 + f_y^2}\,\frac{MTF^2(f_x, f_y)}{NPS(f_x, f_y)} \tag{2.37}$$

Arranging terms from Eq. (2.35) and combining with the $FIM_{\hat{N}\ll G}$ formulation of Eq. (2.18) yields a relationship between image quality and registration performance. For example, in a scenario where the two images are produced by the same imaging system (equivalent MTF), the $FIM_{\hat{N}\ll G}$ in Eq. (2.18) becomes:

$$FIM_{\hat{N}\ll G} = 4\pi A \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix} \tag{2.38a}$$

$$\text{where} \quad \gamma_{ij} = \iint_{-f_{Nyq}}^{f_{Nyq}} \frac{f_i f_j}{\sqrt{f_x^2 + f_y^2}} \frac{NEQ_1 NEQ_2}{NEQ_1 + NEQ_2} G_{Obj}\, df_x df_y \tag{2.38b}$$

where G_{obj} refers to the power spectrum of the object rather than the image of the object (equal to G divided by MTF^2). In this form, we see that registration performance is dependent on high-frequency weighting [carried by the $f_i f_j/\sqrt{f_x^2 + f_y^2}$ term] of the object, in turn modified by the effective NEQ term.

Alternatively, the DQE describes the performance in terms of the dose, MTF, and NPS:

$$DQE(f_x, f_y) = \frac{\pi\sqrt{f_x^2 + f_y^2}}{m\bar{q}} \frac{MTF^2(f_x, f_y)}{NPS(f_x, f_y)} \tag{2.39}$$

Therefore, considering the example of two images produced by the same system (equivalent MTF), rearranging Eq. (2.39) allows the $FIM_{\hat{N}\ll G}$ of Eq. (2.18) to be written as a function of DQE:

$$FIM_{\hat{N}\ll G} = 4\pi A \frac{(m\bar{q})_1 (m\bar{q})_2}{(m\bar{q})_1 + (m\bar{q})_2} \begin{bmatrix} \gamma_{xx} & \gamma_{xy} \\ \gamma_{xy} & \gamma_{yy} \end{bmatrix} \tag{2.40a}$$

$$\text{where} \quad \gamma_{ij} = \iint_{-f_{Nyq}}^{f_{Nyq}} \frac{f_i f_j}{\sqrt{f_x^2 + f_y^2}} DQE \cdot G_{Obj}\, df_x df_y \tag{2.40b}$$

Examination of the RMSE estimate in Eq. (2.33) similarly elucidates the dependence of registration accuracy on spatial resolution characteristics, particularly for the case of maximizing cross correlation. Reduced system MTF (via system blur and/or coarser voxel size) carries the benefit of reduced noise but also reduces the strength of image gradients via H. Therefore, the lower bound on registration accuracy follows a non-monotonic dependence on spatial resolution, suggesting an optimal resolution (alternatively, an optimal post-processing filter) that balances the tradeoffs between noise and gradient strength.

2.3 Experimental Methods

2.3.1 Formation of Test Images

Experiments were conducted based on two digitally simulated axial CT images: (i) a soft-tissue model and (ii) an anthropomorphic head phantom. The soft-tissue model was based on a power-law noise distribution with frequency content as common in statistical modeling of anatomical "clutter" [19], [59], [60] and defined in Sec 1.2.2. A value of $\beta = 3$ has been shown

to model a stochastic arrangement of self-similar, soft-tissue anatomy [59]. A realization of such power-law distribution in 3D was generated and taken as ground truth soft-tissue anatomical structure for CT simulation. For the head image, ground truth was measured from a high-quality CT scan of an anthropomorphic head phantom (The Phantom Laboratory, Greenwich, NY) with soft-tissue manually segmented [42] and set to a constant value of 40 HU.

Simulated CT images of these two models were computed over a broad range in dose by digitally forward-projecting the ground truth images, scaling the fluence in proportion to dose, and adding Poisson noise in proportion to $1/\sqrt{(1+SPR)\times \text{dose}}$, where nominal values for scatter-to-primary ratio (SPR) were chosen: $SPR = 2$ for the soft-tissue image and $SPR = 9$ for the higher attenuation head image [53]. In each case, dose is specified in terms of the x-ray tube current-time product (mAs), which is proportional to absorbed dose via the fluence per unit exposure (q/X), exposure per mAs (X/mAs), and exposure-to-dose conversion (f-factor, cGy/X), all of which are constant for a fixed beam energy (in these studies, a 100 kV spectrum computed using the SPEKTR x-ray simulation toolkit [61])

Each image was simulated from $m = 720$ forward projections over 360°. The fluence was scaled according to total x-ray tube output (mAs) at a beam energy of 100 kV. Images were reconstructed by filtered backprojection, and central 2D axial slices (241×241 pixels for the soft-tissue model and 485×390 for the head) at 0.5 mm × 0.5 mm voxel size were extracted. Example soft-tissue and head images at various dose levels are shown in Fig. 2.1. A total of 22 independent image realizations were generated for each phantom and dose level, each taken as input to the registration process, described below.

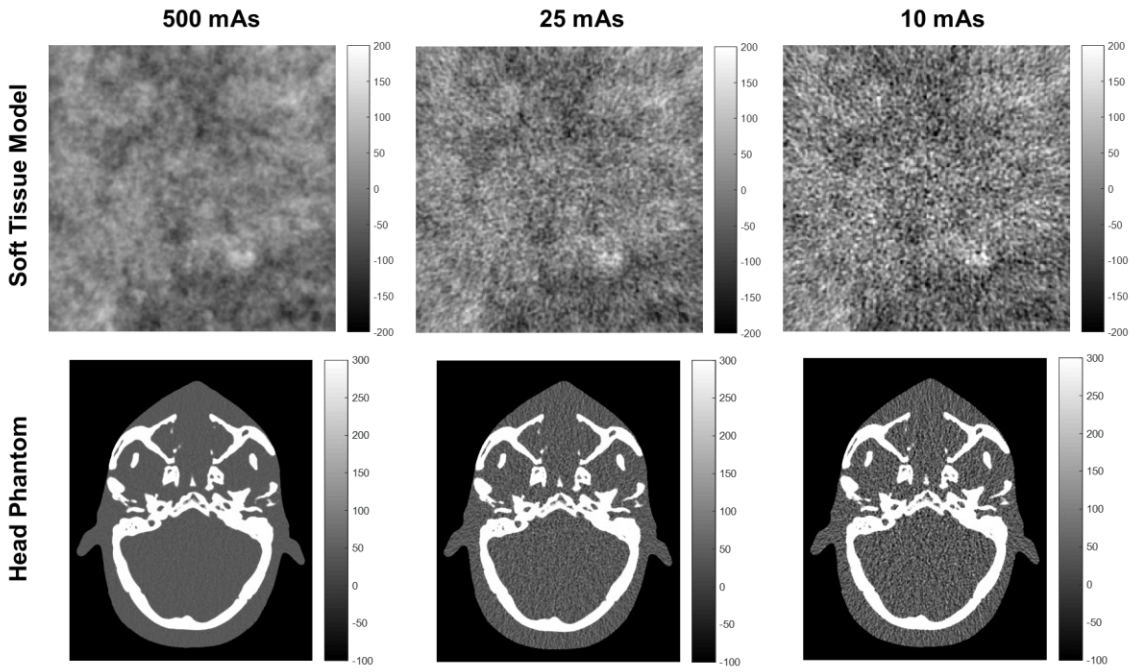


Figure 2.1: Example images of the soft-tissue model (top) and anthropomorphic head phantom (bottom) at various levels of dose (mAs). Figure adapted with permission of the publisher from [44].

2.3.2 Registration Methods and Similarity Metrics

Experiments were performed using three categories of registration (translation-only): (i) The first category involved intensity interpolation, which optimizes a similarity metric over $\widehat{\boldsymbol{\theta}}$ and resamples the image at each iteration under the specified interpolation model (here, cubic B-spline [62]). This was accomplished with SimpleITK [63] using 3 similarity metrics: (a) MSD, (b) Mattes mutual information (MMI) [32] (50 bins), and (c) joint histogram mutual information (JMI) [64] (50 bins, 1.5 σ). (ii) The second category involved metric interpolation, which was computed by evaluating the maximum of a parabolic fit to the normalized cross correlation (NCC-Fit) metric at the pixel-shift peak location and its surrounding 8 pixel-shift neighbors. (iii) Finally, a Fourier-based method of phase correlation (PC) [65] was evaluated.

For a pair of images (generated as in Section 2.3.1), registration was performed using each of these methods after introducing a known shift $\boldsymbol{\theta}$. Prior to inducing the shift, a constant intensity (equal to the mean over the image edges) was subtracted from both images to reduce the effect of zero-padding used for the image transformations.

2.3.3 Performance Evaluation

Registration performance was evaluated in terms of the root-mean-square error (RMSE) of the translation estimate $\widehat{\boldsymbol{\theta}}$. For each of the categories described above, the RMSE was compared to the two forms of lower bound derived in Section III: the CRLB [Eq. (2.12)] and the $\text{CRLB}_{\widehat{N}\ll G}$ [Eq. (2.18)]. The CRLB for an unbiased estimator can be written in terms of RMSE as:

$$RMSE \geq \sqrt{\text{trace}(C_{LB})} = \sqrt{\text{trace}(FIM^{-1})} \tag{2.41}$$

To estimate the power spectra required to calculate the CRLB, the NPS at each mAs (i.e., N_i) was computed by averaging periodograms from a set of 20 instances of simulated noisy images with the mean image subtracted. In computing G, since we only have access to noisy images in the context of Eqs. (2.1) and (2.2), we do not truly know g. In this work, g was formed by computing the mean over 20 images simulated at 500 mAs. We then computed G using the 2D Welch periodogram method [66] using 16 windows (4 increments in each dimension) with an overlap ratio 0.5 and a Hamming window to reduce spectral leakage. While this method is suitable to image simulation, it may not be practical to acquire many instances of an image to compute g. Other methods for approximating G (not investigated in this chapter) include (i) computing the power spectrum from a low-noise (e.g., pre-operative) image

assuming minimal noise, (ii) computing the image power spectrum and subtracting a model estimation of the NPS, or (iii) using a model estimation of the signal power spectrum (e.g., power-law model as fairly common in describing tissue parenchyma [19], [59], [60]). RMSE was analyzed as a function dose (proportional to mAs) and total image variance $\sigma_1{}^2 + \sigma_2{}^2$ (computed by integrating $N_1 + N_2$).

We further evaluated each registration method in terms of the statistical registration efficiency (denoted SRE), defined as:

$$SRE = \frac{\text{trace}(C_{LB})}{(RMSE)^2} \tag{2.42}$$

Written this way, the SRE is bounded ($SRE < 1$) and describes the ratio of the CRLB to the measured mean squared error performance. The SRE was evaluated as a function of dose for each category of registration and similarity metric mentioned above.

2.3.4 Registration Cases

2.3.4.1 Registration of images at equivalent dose (homoscedastic)

Analysis was first performed for image registration in which the noise characteristics of both images were equivalent. Each registration followed the method in Section 2.3.2, with a known shift of $\boldsymbol{\theta} =$ [1.2 pix, 1.2 pix] introduced to the moving image. 231 (i.e., 22 choose 2) registrations were performed between the 22 image realizations formed at the same dose level. A total of 13 dose levels were considered with mAs ranging over 3 orders of magnitude (0.5–500 mAs).

2.3.4.2 Registration of a high-dose to a low-dose image (heteroscedastic)

A common scenario in image-guided interventions was simulated in which a high-dose (i.e., high quality) preoperative image is registered to a low-dose (i.e., lower quality) intraoperative image. The experiment of Sec. 2.3.4.1 was repeated using the MSD similarity metric, considering a fixed dose for the fixed image and varying the dose for the higher-dose moving image. Performance was evaluated in terms of RMSE as a function of the total noise magnitude in the registered images.

2.3.4.3 Effect of image blur on registration performance

We further examined the effect of image blur on image registration performance. Factors affecting blur [described by the image quality model leading to Eq. (2.36)] include the imaging system configuration (e.g., x-ray focal spot size, detector pixel size, and system geometry), reconstruction method (filter kernel), and optional post-processing and/or interpolation filters. The derivation in Sec. 2.2.3.2, leading to the RMSE estimate with postprocessing effects in Eq. (2.33), exposed the non-trivial relationship between system blur and registration performance, suggesting an optimum tradeoff between high-frequency noise magnitude and image signal (i.e., gradient) power. To investigate the effect, the experiment of Sec. 2.3.4.1 was repeated with an additional post-processing Gaussian blur kernel of width σ_b ranging from 0.5 to 7 pixels applied to both images. Results were compared to theoretical predictions for optimal Gaussian blurring using Eq. (2.33).

2.4 Results

2.4.1 Registration Accuracy: Homoscedastic Images

Figure 2.2 shows the performance for the various categories of registration: Fig. (2.2A) metrics MSD, MMI, and JMI; and Fig. (2.2B) methods NCC-Fit and PC — each in comparison to the theoretical lower bound predicted by CRLB [(Eq. 2.12)] and $\mathrm{CRLB}_{\widehat{N}\ll G}$ [(Eq. 2.18)]. In Figure 2.2A, registration performance is seen to improve (i.e., RMSE decreases) with dose for each of the interpolation-based similarity metrics. Each metric performs equivalently at high dose, and each exhibits a low-dose threshold below which registration fails, with MSD demonstrating the strongest robustness to noise and JMI performing the worst. The threshold reflects the noise level at which point estimation errors lie outside of the main lobe of the optimization search space (causing a "failed registration") and leading to arbitrarily large registration errors. The theoretical lower bound predictions appear to be optimistic — i.e., none of the methods achieve the lower bound. However, the overall trend with dose is similar, and the estimators (MSD and MMI in particular) adhere to the trend with similar slope across the broad range of dose levels.

Figure 2.2B shows the same for the NCC-Fit and PC registration methods. (The CRLB and $\mathrm{CRLB}_{\widehat{N}\ll G}$ curves are the same as in Fig. 2.2A.) Interestingly, these methods do not exhibit a low-dose threshold for registration failure, instead following the general trend of the CRLB. The NCC-Fit method appears more robust against noise than PC and follows the CRLB down

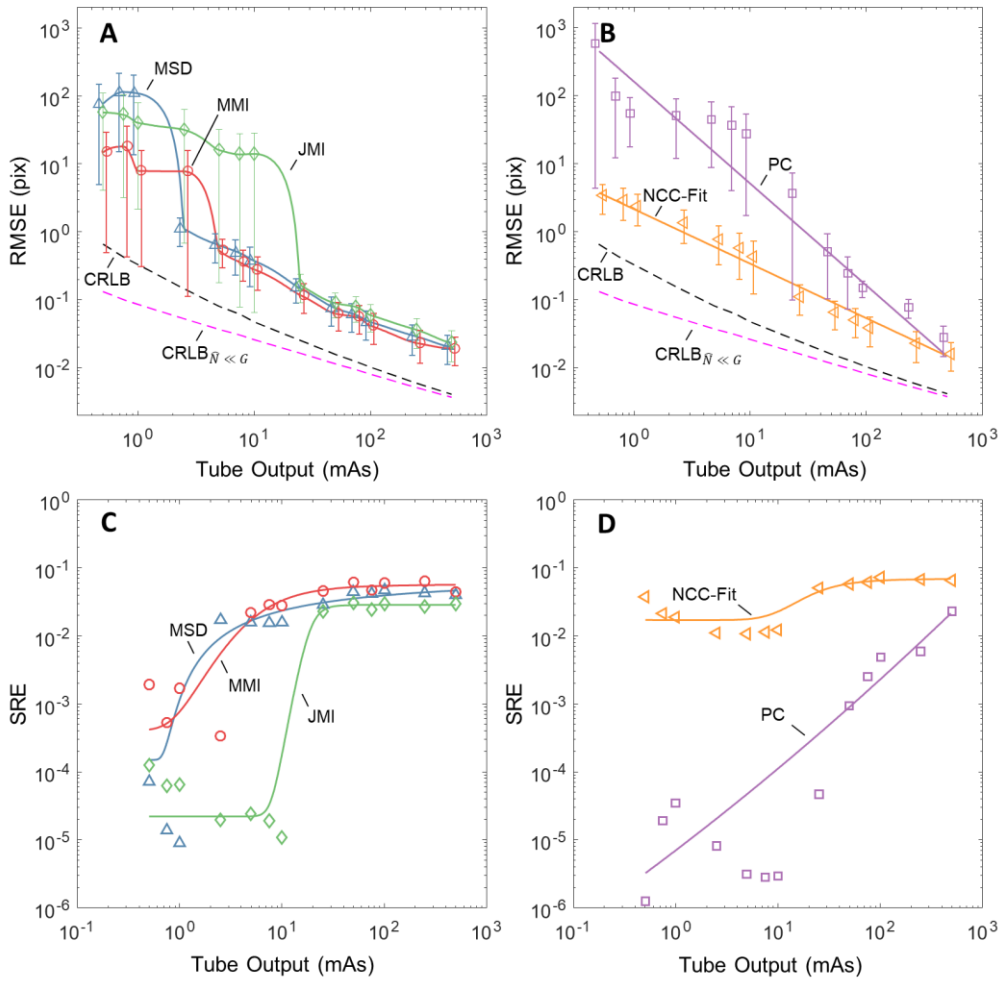


Figure 2.2: Effect of dose on registration performance for the "equal-dose" case (i.e., images with equivalent noise characteristics). Each case is for the soft-tissue images in Figure 2.1. The dashed curves in (A) and (B) mark the lower-bound in registration accuracy predicted by the CRLB and $CRLB_{\bar{N} \ll G}$. (A) RMSE for intensity-interpolation registration using the MSD, MMI, and JMI similarity metrics. (B) RMSE for the NCC-fit and PC registration methods. (C) SRE versus dose for the MSD, MMI, and JMI metrics. (D) SRE for the NCC-fit and PC methods. Figure adapted with permission of the publisher from [44].

to the lowest dose levels investigated. This behavior is attributed in part to the brute-force sampling of NCC-Fit (thus avoiding local minima) and in part to the highly constrained search of NCC-Fit (evaluated only at pixel shifts within a fixed distance from solution), which avoids the large RMSE registration failure threshold effect. The PC registration method exhibits poorer robustness to noise (higher negative slope) and steady degradation to registration failure with reduced dose.

Figures 2.2C and 2.2D illustrate the extent to which various methods achieve the CRLB by evaluating the SRE versus dose. Figure 2.2C shows the SRE for the intensity-interpolation metrics (MSD, MMI, and JMI), showing that each approaches SRE ~0.04 at high dose, but efficiency falls by more than an order of magnitude at the low-dose threshold identified in Figure 2.2A. Figure 2.2D shows that the NCC-Fit method maintains SRE over the entire range of dose investigated (again, likely attributed to the highly constrained search), whereas the PC method shows a steep degradation in SRE with reduced dose.

At high dose, each metric and method achieved SRE of just ~0.04 for the soft-tissue phantom (and ~0.11 for the anthropomorphic head, not shown for brevity). When considering this fairly low level of SRE, it should first be noted that the CRLB is generally not guaranteed to be obtainable. Moreover, even when it is obtainable, only a selection of estimators may be able to achieve the bound, and often only asymptotically — i.e., in a manner that requires larger and larger data size to achieve the bound. Since the images used in this study were relatively small and optimal estimators were not examined [e.g., optimal filtering to minimize Eq. (2.33)], we do not expect the result to achieve an SRE of 1. Given these points, it is also important to note that errors in power spectrum estimation (particularly in the high frequency region) may have contributed to an optimistic CRLB. The use of physics-guided models for the power spectra along with robust estimation methods may provide a more accurate estimation of the CRLB (presented in Chapter 3).

Examining Eq. (2.40), we see that in the homoscedastic (equal-dose) case, registration error in the strong signal approximation is proportional to $1/\sqrt{\text{dose}}$, which is evident in the slope of the $\text{CRLB}_{\widehat{N} \ll G}$ curve (Figs. 2.2A and 2.2B). For the soft-tissue image case (dominated by low-frequency signal power), the high-signal approximation appears to hold well only at

high dose (i.e., low noise). For the head image (which exhibits a greater proportion of mid- and high-frequency signal power), the approximation holds within 15% of CRLB over a broader dose range — down to ~2.5 mAs. In the lower dose range, disagreement between CRLB with $\mathrm{CRLB}_{\hat{N}\ll G}$ arises due to increased influence of the $N_1 N_2$ cross-correlation noise term.

2.4.2 Registration Accuracy: Heteroscedastic Images

Figure 2.3 shows MSD registration performance as a function of (total) noise magnitude for the heteroscedastic case in which a low-noise (i.e., higher-dose) image is registered to a noisier (i.e., lower-dose) image. The RMSE is plotted versus total variance ($\sigma_1{}^2 + \sigma_2{}^2$) for the (A) soft-tissue and (B) head phantom images of Fig. 2.1. The CRLB and $\mathrm{CRLB}_{\hat{N}\ll G}$ theoretical lower limits are shown as dashed lines. The color scale on the plot symbols and curve fits refers to the mAs of the lower-dose image: for example, the lower-left of each plot shows the (red) case for which both images were formed at the maximum dose (500 mAs), whereas the upper-right of each plot shows the (black) case for which the lower-dose image was formed at just 0.5 mAs.

Whereas Figs. 2.2A–B demonstrated that RMSE is proportional to $1/\sqrt{\mathrm{dose}}$ and thus $\sqrt{\sigma_1{}^2 + \sigma_2{}^2}$, this simple relationship is lost in the heteroscedastic case shown in Fig. 2.3. In Fig. 2.3A, we observe a highly non-linear dependence on the total noise; however, this non-linearity is predicted very well by the CRLB. The $\mathrm{CRLB}_{\hat{N}\ll G}$ approximation, however, only

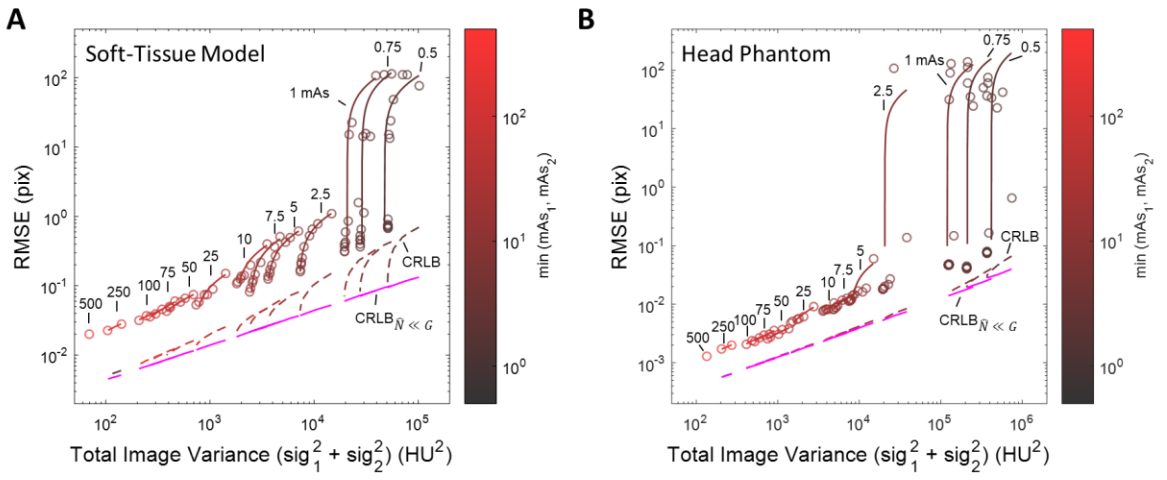


Figure 2.3: Registration performance (using MSD) versus total image noise for the heteroscedastic case: (A) soft tissue image and (B) head phantom image. Each circle represents the RMSE for a specific I_1, I_2 dose level combination, with connected circles of the same color indicating the same mAs for the low-dose image. The colorscale and labels denote the mAs for the lower-dose image. The CRLB (dashed) and CRLB$_{\hat{N} \ll G}$ (magenta) formulations are also plotted. Figure adapted with permission of the publisher from [44].

describes the effect of relative dose through the $(m\bar{q})_1(m\bar{q})_2/[(m\bar{q})_1 + (m\bar{q})_2]$ term, appropriate only at low-noise (high-dose) conditions for the soft-tissue model.

Figure 2.3B shows similar trends in registration performance for the head phantom image. In this case, however, the image exhibits sufficiently large high-frequency signal power, such that $N_1 N_2 \ll GN_1 + GN_2$. As a result, the CRLB is roughly proportional to $\sqrt{{\sigma_1}^2 + {\sigma_2}^2}$ and agrees with the CRLB$_{\hat{N} \ll G}$ approximation over a much broader range of dose.

2.4.3 Registration Accuracy: Effect of Image Blur

Figure 2.4A shows the registration performance (RMSE) for the heteroscedastic soft-tissue case (as in Fig. 2.3A), comparing the RMSE achieved by the MSD method (which can be shown to be nearly equivalent to maximizing cross-correlation) with that predicted by Eq. (2.32). We immediately see that while Eq. (2.32) somewhat underestimates the magnitude

of RMSE, it trends well with the measured dependence of registration performance on dose, yielding a correlation coefficient of R = 0.988 between the predicted and measured RMSE in the higher dose region [where the Taylor approximation in Eq. (2.32) is appropriate].

Figure 2.4B summarizes the findings of optimal post-processing blur for registration of the soft-tissue image at various dose levels. Each curve represents the RMSE (at a given dose level) as a function of blur width (σ_b). The blue star marks the measured minimum in RMSE (i.e., optimal blur), and the magenta circle marks the theoretical minimum as predicted by minimizing Eq. (2.33) with respect to H_{CC}. As expected, post-processing blur is most beneficial under high-noise (low-dose) conditions (black curves). On the other hand, for low-noise (high-dose) conditions (red curves), excessive blur is seen to degrade registration performance. The measured and predicted values for optimal blur agree with this trend and match fairly well across a broad range of dose, further validating the model of Eq. (2.33) as a figure of merit for registration.

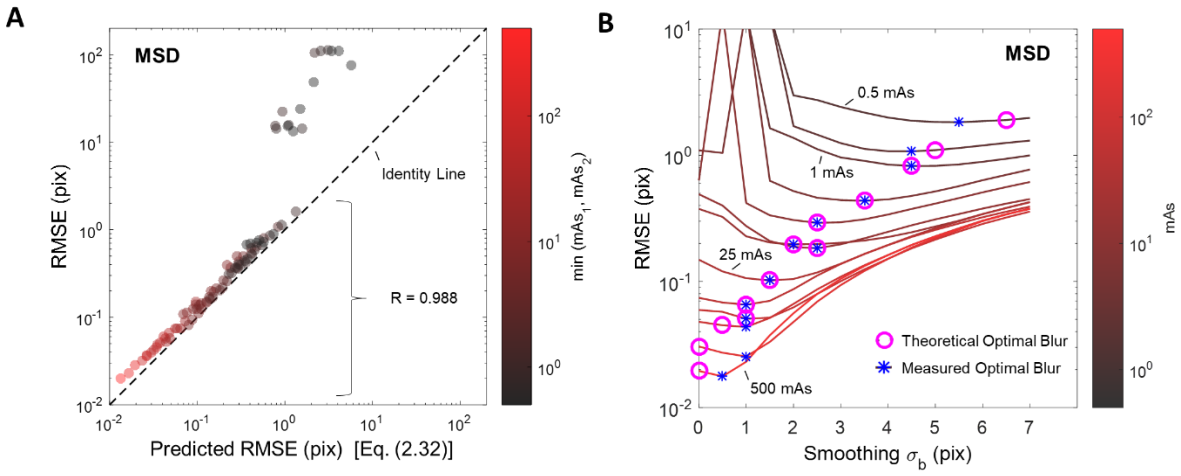


Figure 2.4: (A) Error in soft-tissue image registration compared to the performance predicted by Equation (2.32). (B) Registration performance as a function of post-processing blur at various dose levels. The results pertain to the MSD registration method, and dose reflected in the mAs colorscale. For each curve, the magenta circle represents the predicted optimal blur level, and the blue star represents the measured optimal blur. Figure adapted with permission of the publisher from [44].

Figure 2.5 further investigates the predicted benefit of post-processing blur on registration performance for several similarity metrics. The data correspond to the soft-tissue image model, homoscedastic image registration, and theoretically optimal Gaussian blur (OGB) derived by minimization of Eq. (2.33) with respect to H_{CC}. The SRE is plotted versus dose, and we observe that application of an optimal Gaussian blur maintains optimality (again at a level of SRE ~0.04) across the range of dose levels investigated for MMI and MSD, bearing in mind that Eq. (2.33) applies directly only to cross-correlation based metrics, e.g., MSD. Close inspection of Figure 2.5 suggests a slight increase in SRE at the lower dose levels — a somewhat surprising result that is in agreement with the theoretical prediction. The increase in efficiency is because Gaussian blur is not a truly optimal filter as described by H_{HT} of Eq. (2.34) in minimizing Eq. (2.33); however, with respect to Gaussian filters, the OGB may more closely approximate H_{HT} under low-dose / high-noise conditions, leading to the increase in SRE.

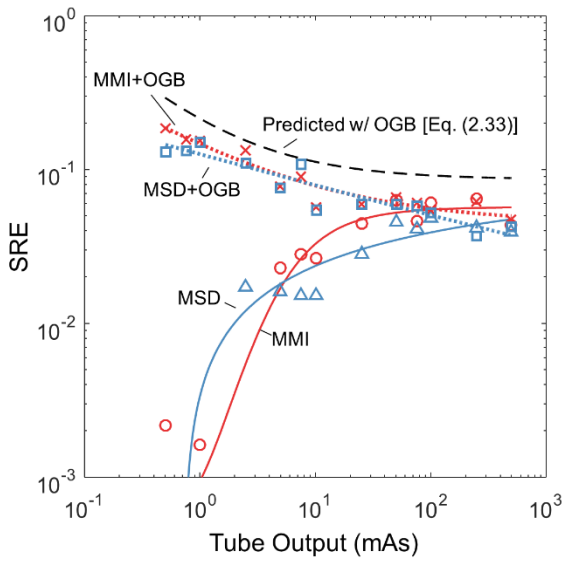


Figure 2.5: SRE evaluated as function of dose for MSD (blue) and MMI (red) with and without optimal Gaussian blur (OGB). The predicted SRE (with OGB) is shown as the black dashed line, demonstrating a similar dose dependence as the measurements with optimal blur. Figure adapted with permission of the publisher from [44].

2.5 Conclusion

As discussed in Sec. 1.2.2, an image is always acquired for a task and evaluation of imaging system performance should be with respect to that task. For many scenarios in image-guided interventions, the task may relate to registration of image information more so than to visualization. The framework described above provides a means by which to evaluate an imaging system with respect to registration performance, providing a basis for assessing the performance of various registration methods and selecting optimal image acquisition and reconstruction techniques.

As a first step in establishing this framework, we presented theoretical lower bounds for registration performance and investigated several sub-pixel estimators as a function of dose and noise magnitude. Following this analysis, we examined the registration method of maximizing cross-correlation to analyze the effect of spatial resolution on registration performance, thereby deriving an RMSE figure of merit, Eq. (2.33). The model was shown to agree well with measurements of registration accuracy for various choices of post-processing linear filters (blur), providing understanding beyond the basic information-theoretic data processing effect [in which linear filters have no effect on SNR in Eq. (2.15)] and guiding selection of optimal filters that extend registration performance to lower dose levels.

In the derivation of the FIM of Eq. (2.12), we assumed stationary Gaussian characteristics of the signal and noise. As discussed in [67], the Gaussian assumption is not a particularly strong requirement, since (by the Central Limit Theorem) even when the signal is not Gaussian the Fourier coefficients of the signal will tend toward a Gaussian distribution as the number of image samples increases. Nonetheless, further investigation is warranted to more fully account

for nonstationary MTF and NPS characteristics in CT images [68] as well as non-Gaussian characteristics. However, in line with similar approximations for analysis of detectability index [11], we see in Sec. 2.4.2 that despite the approximation Eq. (2.12) proves a useful predictor of performance trends, even in the case of a high-contrast (head) image that exhibits highly non-Gaussian, non-stationary characteristics.

The current work examined the simple case of 2D translation-only intra-modality registration. Extension of the formulation to 3D translation is straightforward, as shown in Sec. 2.2.2.1. Multi-modality registration is not considered in the current work, as significant modification to the statistical model would be required to capture the mismatch in the image content. In future work, because image quality models (such as [53], [40]) permit analysis of the spatially varying (i.e., nonstationary) *local* MTF and NPS, the analysis shown above can be similarly extended to description of local registration accuracy in regions differing in image quality. Using these local approximations to extend the analysis to deformable image registration is an exciting possibility (examined in Chapter 5).

Throughout this work is the assumption that the registration method is unbiased (intrinsic to $C_{LB} = FIM^{-1}$). This assumption breaks down at least in part for the NCC-fit registration method [51] due to the parabolic fit and has yet to be rigorously investigated for the other optimization methods. While it is likely that there are small biases in the other methods as well (owing to choice of interpolation, optimizer, etc.), the assumption of unbiased estimators appears to be reasonable, as the observed RMSE was dominated by the variance term (rather than the bias), and the experiments demonstrated similar trends as the CRLB (whereas a plateau effect would likely be observed in the low-noise region for a system dominated by bias).

The FIM provides a framework that is independent of the particular registration method — whether biased or unbiased. We extended previous work in CRLB estimation in such problems by generalizing to 2D and 3D, allowing for disparate noise in the I_1 and I_2 images, and including image blur as well as noise correlation, demonstrating results beyond a simple approach of AWGN (which is a poor approximation to noise in CT / CBCT). The resulting analytical framework leverages well-established models describing image quality in CT/CBCT [6], [7], [10] as in Eq. (2.36) and is consistent with the theme of task-based imaging performance. With respect to image-guided interventions, the analysis provides a new framework for understanding the performance of imaging systems with respect to the task of image registration. In the next chapter, we extend this framework to model the confounding effect of soft-tissue deformation when registering rigid anatomy.

Chapter 3: A Statistical Model for the Influence of Soft-Tissue Deformation on Rigid Image Registration Performance

3.1 Introduction

The statistical framework discussed in the previous chapter provides insight on the effects of image quality (viz., dose, quantum noise, and spatial resolution) on image registration performance — providing a basis by which the image acquisition and postprocessing techniques can be optimized according to the task of registration. However, in practice the underlying assumptions are in part broken when structures in g are subject to deformation between I_1 and I_2, suggesting a disparity in the true underlying signal (g). For example, anatomy presenting in medical images often consists of rigid (bone) and deformable (soft tissue) components. In such a scenario, despite soft-tissue deformation, bone anatomy still provides salient structure suitable to accurate rigid registration. By considering the rigid anatomy to be the "true" underlying signal g — corresponding to scenarios in which the rigid

anatomy is the structure of interest (e.g., orthopedic surgery) — we may construct a model in which non-rigid, soft-tissue structures are considered as a confounding noise source with respect to the task of rigid registration.

This approach is analogous to approaches drawn from SDT (Sec. 1.2.2) in which background anatomical "clutter" is considered as a confounding noise source with respect to the task of detection [69]–[71]. As discussed in Sec. 1.2.2, such SDT frameworks have provided an important basis for imaging system optimization (e.g., in flat-panel detectors [72] and cone-beam CT [73], [74]), and an important aspect of such models is a generalization in which not only quantum noise is considered as a confounding influence on detection, but so is *any* fluctuation in the image that is not associated with the stimulus (e.g., background lung or breast parenchyma) [75], [76]. Such generalization of the visual detection process is clearly an abstraction, since background anatomy is not a random process, but it has provided a useful analytical basis for guiding important aspects of imaging system design — e.g., evaluating gains in detectability of a particular stimulus in projection, tomosynthesis, and fully 3D tomographic imaging and the point beyond which detection is not improved by increasing dose.

In a similar manner, rigid registration of a rigid (bone) structure can be confounded by nearby soft-tissue deformation acting as "noise" in the similarity metric calculation. In this chapter, we will extend the statistical model of Chapter 2 — namely the CRLB of Eq. (2.12) and the RMSE estimate of Eq. (2.33) — to incorporate soft-tissue deformation as a source of noise in rigid image registration for two common intraoperative scenarios — 3D-2D registration (e.g., preoperative CT to intraoperative fluoroscopy) and 3D-3D registration (e.g., preoperative CT to intraoperative CBCT) in spine surgery.

The work appearing in this chapter was reported in the following journal paper: (M.D. Ketcha et al., *IEEE Trans. Med. Imag.,* 38(9), 2019) [77].

3.2 Model for Soft-Tissue Deformation

3.2.1 Soft-Tissue Deformation as a Noise Source

We consider two cases of soft-tissue deformation, the first being 3D-2D registration [78], [79] in which a 2D digitally reconstructed radiograph (DRR) is computed from a preoperative 3D CT volume and aligned to an intraoperative radiograph as illustrated in Fig. 1.1. Note that this process corresponds to projection-based 3D-2D registration, not slice-to-volume registration. Throughout this chapter, projection images — radiographs or DRRs — are referred to as 2D, and volumetric images — namely CT — are termed 3D, even with respect to a single slice drawn from a 3D CT volume. In 3D-2D registration scenarios, the impact of soft-tissue deformation on registration of bone anatomy can be large, since thick regions of soft tissue carry a high degree of power in the image (obscuring even bone), and deformations can be large (since the patient moves between 2D and 3D imaging systems). To compensate for this, soft tissue is often thresholded out of the CT image (by intensity threshold) before generating the DRR, presenting an "absence" of soft tissue compared to the radiograph. Since the soft tissue is present in only one image, it acts as an independent additive noise source described by the signal and noise decompositions of I_1 and I_2 in Eqs. (2.1) and (2.2). Therefore, soft tissue can be easily incorporated in the model by modifying the noise-term (n_2, taking the radiograph to be I_2) to contain both quantum noise (q_2) and soft-tissue anatomical

noise (s_2), giving $n_2(x,y) = q_2(x,y) + s_2(x,y)$. With $n_2(x,y)$ defined in this manner, $g(x,y)$ represents just the bone anatomy, and $n_1(x,y)$ represents the quantum noise projected in the DRR.

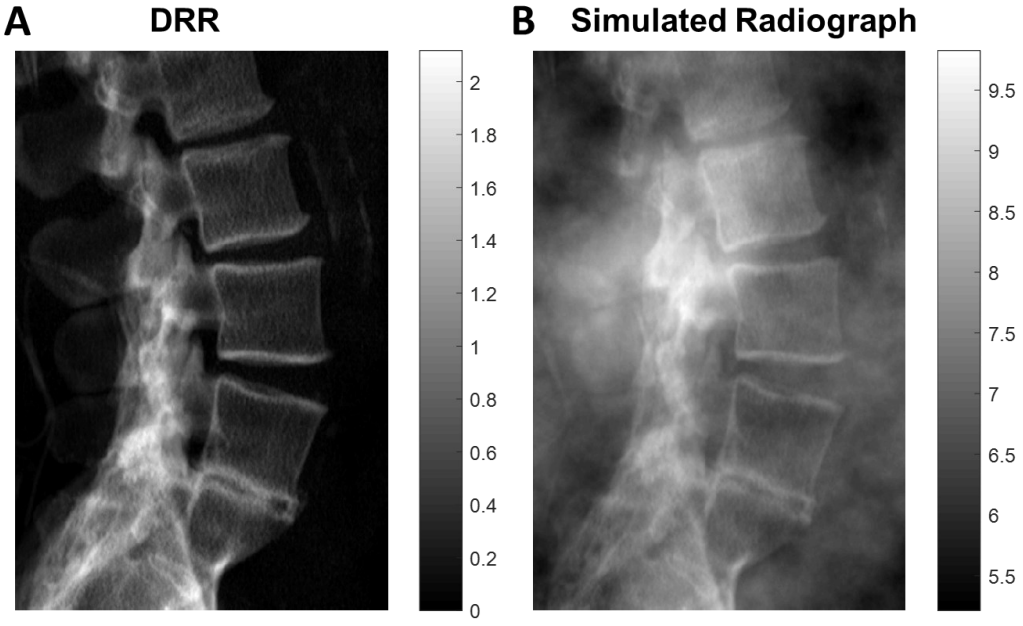


Figure 3.1: 3D-2D registration. (A) Lateral DRR computed from a preoperative 3D CT image thresholded to remove soft tissue. (B) Intraoperative lateral 2D radiograph — in this case, simulated from the DRR in (A) with the addition of power-law soft-tissue anatomical noise. Figure adapted with permission of the publisher from [77].

The second case considers soft-tissue deformation in 3D-3D image registration, starting with the example of registering two axial CT slices, as illustrated in Fig. 3.2. Rigid registration in the presence of soft-tissue deformation can still accurately align bone anatomy ($g(x,y)$), leaving residual misalignment of the deformed soft tissue. From an optimization standpoint, this misalignment of soft-tissue structures (depicted in the colorwash of Figs. 3.2B and 3.2D) diminishes the similarity metric and reduces the quality of the search space, including introduction of false local minima. A problem introduced in modeling deformed soft tissue as noise is that the noise terms in Eqs. (2.1) and (2.2) are assumed independent, which is not the case in this scenario, since soft tissue presenting in one image is just a deformed version of its

manifestation in the other. However, if the deformation is large compared to the correlation length of the gradient image (i.e., high-gradient regions are no longer overlapping), then $s_1(x, y)$ and $s_2(x, y)$ may be approximated as independent. Therefore, both images carry a noise term: $n_i(x, y) = q_i(x, y) + s_i(x, y)$. Note that in the case of no deformation, the soft-tissue function is rightly incorporated in the true signal ($g(x, y) \rightarrow g(x, y) + s(x, y)$) as it contributes positively to the similarity metric.

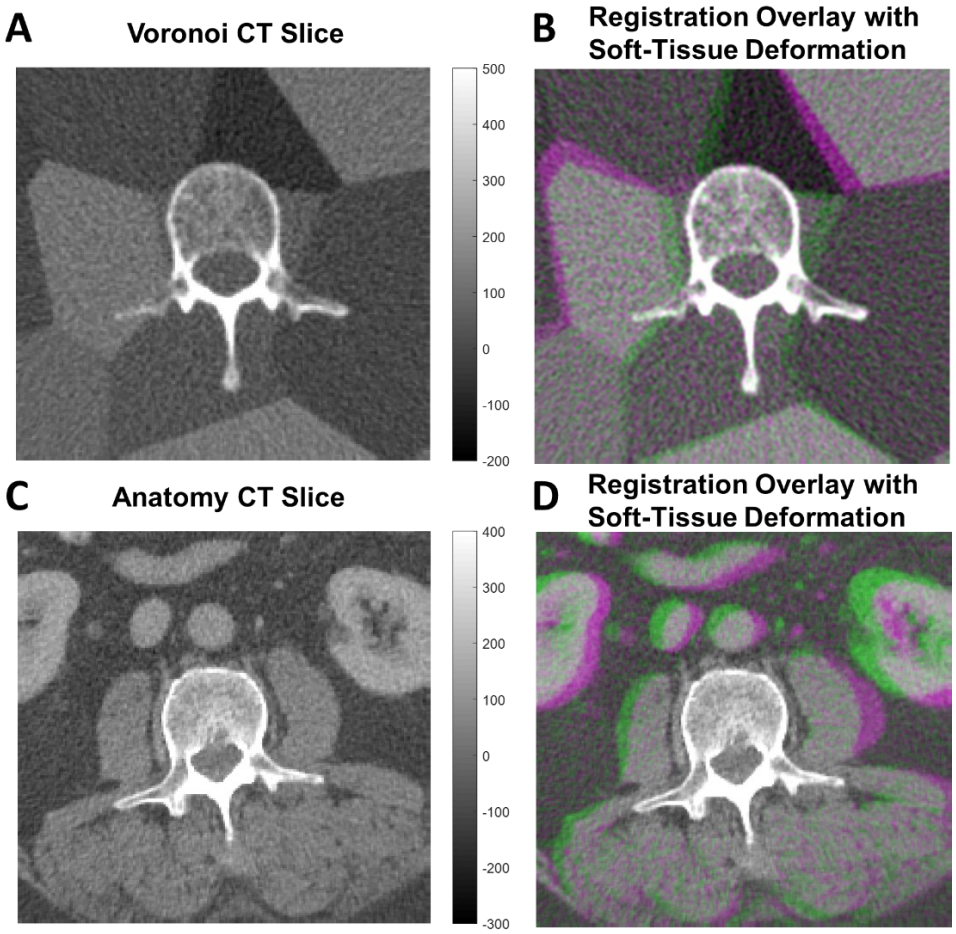


Figure 3.2: 3D-3D registration. (A) Axial CT with a rigid bone (vertebra) and simulated soft-tissue background approximated by a deformable Voronoi distribution of piece-wise constant regions. (B) Colorwash depicting misalignment (green/magenta) of soft tissues following rigid registration. (C) Axial CT image showing real anatomy (abdominal CT). (D) Colorwash depicting misalignment (green/magenta) of soft tissues following rigid registration. Figure adapted with permission of the publisher from [77].

3.2.2 The Soft-Tissue Power Spectrum

To incorporate noise associated with soft-tissue deformation into the figures of merit from Chapter 2, we need a model for the power spectrum of $n_i(x,y)$. We first note that under the assumption that the quantum noise and soft-tissue signals are independent, then the power spectrum of $n_i(x,y) = q_i(x,y) + s_i(x,y)$ is the sum of the two power spectra, giving $N_i(f_x,f_y) = Q_i(f_x,f_y) + S_i(f_x,f_y)$, where Q_i is the quantum NPS and S_i is the soft-tissue power spectrum. As discussed in Sec. 1.2.1, quantum noise in both radiographs and CT images have been well described by models of the quantum NPS, including factors determined by the acquisition technique (e.g., energy and exposure) and imaging system characteristics (e.g., blur, pixel size, and electronic noise) [6], [7], [10]. Furthermore, we saw in Sec. 1.2.2.2 that the power spectrum associated with cluttered scenes (e.g., soft-tissue anatomy overlying structures of interest) have been described from the standpoint of statistical decision theory in terms of a power-law distribution:

$$S_{obj}(f_x,f_y) = \frac{\alpha_S}{f_0^{\beta_S} + f^{\beta_S}} \tag{3.1}$$

where $f = \sqrt{f_x^2 + f_y^2}$ and S_{obj} refers to the power spectrum of the object (cf., the image of the object, which is further attenuated by the MTF^2). Equation (3.1) is similar to the power-law distribution discussed in Sec. 1.2.2, with the parameter α_S scaling in proportion to the tissue contrast and β_S governing the low-frequency extent (generally ranging from 2–4 [19]–[22]); however, Eq. (3.1) further includes f_0 to remove the discontinuity at $f = 0$ (as discussed in [71]), nominally set to be the inverse of the image width.

Sampling from a power-law distribution yields the lumpy background texture seen in Fig. 3.1B, which is appropriate to radiographic and mammographic anatomy. However, the soft-tissue background associated with an axial CT slice tends to follow a piece-wise constant texture. A Voronoi distribution is therefore proposed to simulate axial CT soft-tissue images, with randomly placed seed points and piece-wise constant background defined by intensity values drawn from a uniform distribution over the range of soft-tissue HU values as shown in Fig. 3.2A. While the Voronoi diagram is not a perfect model for solid-organ tissues in tomography (which contain a degree of heterogenous structure owing to density variations, of course) it provides a reasonable first-order approximation. Furthermore, as derived in the next section, the power spectrum for the Voronoi image is shown to be well approximated by a power-law distribution, offering an analytical form similar to previous models of (radiographic) power-law clutter.

3.2.3 Derivation of the Voronoi Power Spectrum

3.2.3.1 Derivation for 1D Voronoi Diagrams

To gain analytical insight on the power spectrum of the piece-wise constant Voronoi image, we begin by considering an analogous 1D case constructed by summing randomly scaled and shifted rect functions:

$$g(t) = \sum_{i=1}^{n} A_i \,\mathrm{rect}\left(\frac{t - t_{0_i}}{T_i}\right) \tag{3.2}$$

where $T \sim \text{Uniform}(T_{min}, T_{max})$ and $E\{A^2\}$ is finite. By utilizing the Fourier pair relating the rect function to the sinc:

$$A \,\text{rect}\left(\frac{t - t_0}{T}\right) \overset{\mathcal{F}}{\Leftrightarrow} AT\text{sinc}(\pi f T)\exp(-i2\pi f t_0) \tag{3.3}$$

the expected power spectrum of $G(f) = \mathcal{F}\{g(t)\}\,\overline{\mathcal{F}\{g(t)\}}$ is:

$$\begin{aligned} E\{G_{1D}(f)\} &= \sum_{i=1}^{n} E\left\{A_i^2 T_i^2 \text{sinc}^2(\pi f T_i)\right\} \\ &= \sum_{i=1}^{n} E\left\{\frac{A_i^2 T_i^2 \sin^2(\pi f T_i)}{\pi^2 f^2 T_i^2}\right\} \\ &= \frac{E\{A^2\}}{\pi^2 f^2}\sum_{i=1}^{n} E\left\{\sin^2(\pi f T_i)\right\} \end{aligned} \tag{3.4}$$

The first equality assumes independence of the summed rect functions, leaving only the summation of the expectations, and the multiplication of complex conjugates cancels the exponential terms. By writing out the sinc, we need only to compute the expectation of the $\sin^2(\cdot)$ term over T, giving:

$$E\{G_{1D}(f)\} = \frac{n\,E\{A^2\}}{\pi^2 f^2}\left(\frac{1}{2} + \frac{\sin(2\pi f T_{min}) - \sin(2\pi f T_{max})}{4\pi f (T_{max} - T_{min})}\right) \tag{3.5}$$

which for large $(T_{max} - T_{min})$, yields the result:

$$E\{G_{1D}(f)\} \approx \frac{1}{2}\frac{n\,E\{A^2\}}{\pi^2 f^2} \tag{3.6}$$

Therefore, in 1D, we see this piece-wise constant function follows a power-law distribution with $\beta_s = 2$.

3.2.3.2 Derivation for 2D Voronoi Diagrams

We extend the derivation to 2D by approximating the Voronoi image as a sum of 2D rects with random rotation (θ):

$$g(x,y) = \sum_{i=1}^{n} A_i \,\mathrm{rect}\left(\frac{x - x_{0_i}}{X_i}, \frac{y - y_{0_i}}{Y_i}; \theta_i\right) \tag{3.7}$$

where $X \sim \mathrm{Uniform}(X_{min}, X_{max})$, $Y \sim \mathrm{Uniform}(Y_{min}, Y_{max})$, $\theta \sim \mathrm{Uniform}(0, 2\pi)$, and $E\{A^2\}$ is finite. As the Fourier transform of a rotated function is simply the rotation of the Fourier transform, we begin by computing the power spectrum of unrotated rect functions and then compute the expectation over θ in Fourier space:

$$\begin{aligned} E\{G_{2D}(f_x, f_y)\} &= \sum_{i=1}^{n} E\{A_i^2 X_i^2 Y_i^2 \,\mathrm{sinc}^2(\pi X_i f_x)\,\mathrm{sinc}^2(\pi Y_i f_y)\} \\ &= E\{A^2\} \sum_{i=1}^{n} E\left\{\frac{\sin^2(\pi X_i f_x)\sin^2(\pi Y_i f_y)}{\pi^4 f_x^2 f_y^2}\right\} \end{aligned} \tag{3.8}$$

Rewriting in polar coordinates and simplifying, we have:

$$E\{G_{2D}(f,\theta)\} = \frac{E\{A^2\}}{\pi^4 f^4} \sum_{i=1}^{n} E\left\{\frac{\sin^2(\pi X_i f\cos(\theta))\sin^2(\pi Y_i f\sin(\theta))}{\cos^2(\theta)\sin^2(\theta)}\right\} \tag{3.9}$$

where the expectation of the inner function is computed over X, Y,and θ. Numerical simulation showed this expectation to closely follow:

$$\begin{aligned} E\left\{\frac{\sin^2(\pi X_i f\cos(\theta))\sin^2(\pi Y_i f\sin(\theta))}{\cos^2(\theta)\sin^2(\theta)}\right\} &\approx \pi f\left(\frac{X_{max} + X_{min} + Y_{max} + Y_{min}}{4}\right) \\ &= \pi f\, \mu_{XY} \end{aligned} \tag{3.10}$$

for large values of $(X_{max} - X_{\min})$ and $(Y_{max} - Y_{\min})$, where we simplify notation by using

μ_{XY} to refer to the mean over the uniform random variable parameters for the rect widths, giving:

$$E\{G_{2D}(f_x, f_y)\} \approx \frac{nE\{A^2\}}{\pi^3 f^3}\mu_{XY} \tag{3.11}$$

We see that Eq. (3.11) again follows a power law distribution, this time with $\beta_s = 3$. In this way, the Voronoi image yields a random image model that is visually similar to the piece-wise constant background of soft tissue presenting in axial CT and has a power spectrum in line with the models derived previously for detection of a signal against a lumpy background with $\beta_s = 3$. Note that β_s is independent of the number of rect functions (n) and widths, implying that a Voronoi image of any density of seed points follows a power law distribution with $\beta_s = 3$. This point is confirmed by the power spectra measured for randomly generated Voronoi images described in Sec. 3.2.2.

3.2.3.3 Derivation for 3D Voronoi Diagrams

Extending the derivation to 3D rect functions begins by incorporating spherical rotations so that:

$$g(x, y, z) = \sum_{i=1}^{n} A_i \operatorname{rect}\left(\frac{x - x_{0_i}}{X_i}, \frac{y - y_{0_i}}{Y_i}, \frac{z - z_{0_i}}{Z_i}; \theta_i, \varphi_i\right) \tag{3.12}$$

The distributions of the random variables are identical to the 2D case, where now $Z \sim \text{Uniform}(Z_{min}, Z_{max})$ and φ ranges from 0 to 2π and follows cumulative distribution

function $F(\varphi) = (1 - \cos(\varphi))/2$ so that the spherical rotations uniformly sample the sphere. Similarly:

$$\begin{aligned} E\{G_{3D}(f_x, f_y, f_z)\} &= \sum_{i=1}^{n} E\{A_i^2 X_i^2 Y_i^2 \, Z_i^2 \mathrm{sinc}^2(\pi X_i f_x) \, \mathrm{sinc}^2(\pi Y_i f_y) \mathrm{sinc}^2(\pi Z_i f_z)\} \\ &= E\{A^2\} \sum_{i=1}^{n} E\left\{ \frac{\sin^2(\pi X_i f_x) \, \sin^2(\pi Y_i f_y) \sin^2(\pi Z_i f_z)}{\pi^6 f_x^2 f_y^2 f_z^2} \right\} \end{aligned} \tag{3.13}$$

By converting to spherical coordinates, numerical simulation shows the expectation to closely follow the form:

$$E\{G(f_x, f_y, f_z)\} \approx 2 \frac{nE\{A^2\}}{\pi^4 f^4} \mu_{XYZ}^2 \tag{3.14}$$

where $f = \sqrt{f_x^2 + f_y^2 + f_z^2}$ and μ_{XYZ} is mean over the six uniform distribution width parameters. In this form, we observe that the 3D Voronoi distribution also follows a power-law distribution, though now with $\beta_s = 4$.

3.2.4 Robust Registration Methods

In Sec. 2.2.3.2, we presented a method to optimize registration performance by minimizing the RMSE estimate, Eq. (2.33), with respect to the Gaussian blur filter, $H_{CC}(f_x, f_y; \sigma_b)$ of Eq. (2.35). While the method is particularly useful for registration methods utilizing the NCC loss function, experimental studies investigating 3D-2D registration [78], [80] have shown that gradient-based metrics provide more robust registration performance

compared CC-based methods. Therefore, we will further investigate alternative similarity metrics such as gradient correlation (GC), defined as:

$$GC(\hat{u},\hat{v}) = \left(\frac{\partial I_1}{\partial x} \otimes \frac{\partial I_2}{\partial x}\right)(\hat{u},\hat{v}) + \left(\frac{\partial I_1}{\partial y} \otimes \frac{\partial I_2}{\partial y}\right)(\hat{u},\hat{v}) \tag{3.15}$$

Equation (3.15) shows that GC is the sum of the cross-correlation of the partial derivative images ($\partial I_i/\partial x\,, \partial I_i/\partial y$). These partial derivative images are typically computed by convolving the images with spatial derivative filters $h_x(x,y), h_y(x,y)$ (e.g., Sobel, derivative of a Gaussian, etc.), thus we rewrite (3.15) as:

$$\begin{aligned} GC(\hat{u},\hat{v}) &= (h_x * I_1)\otimes(h_x * I_2) + \big(h_y * I_1\big)\otimes\big(h_y * I_2\big) \\ &= (h_x \otimes h_x) * (I_1 \otimes I_2) + \big(h_y \otimes h_y\big) * (I_1 \otimes I_2) \\ &= \big(h_x \otimes h_x + h_y \otimes h_y\big) * (I_1 \otimes I_2) \\ &\overset{\mathcal{F}}{\Leftrightarrow} H_{GC} \cdot \mathcal{F}\{I_1 \otimes I_2\} = H_{GC} \cdot \mathcal{F}\{I_1\} \cdot \overline{\mathcal{F}\{I_2\}} \end{aligned} \tag{3.16}$$

where the Fourier transform in the last line shows that GC can be computed by filtering the CC function $(I_1 \otimes I_2)$ with the function $H_{GC}\big(f_x\,, f_y\big)$. When $h_x(x,y)$ and $h_y(x,y)$ are the derivative of Gaussian spatial derivative filters, we have:

$$H_{GC}\big(f_x\,, f_y; \sigma_b\big) = \big(f_x^2 + f_y^2\big)\, e^{-4\pi^2(f_x^2+f_y^2)\sigma_b^2} \tag{3.17}$$

which can be used with Eq. (2.33) to optimize registration performance for the GC similarity metric. Furthermore, a more general form for the *n*-th derivative of a Gaussian spatial filter is:

$$H_{Gn}\big(f_x\,, f_y; n, \sigma_b\big) = \big(f_x^2 + f_y^2\big)^n\, e^{-4\pi^2(f_x^2+f_y^2)\sigma_b^2} \tag{3.18}$$

allowing one to model the performance of higher-order gradient similarity metrics (referred to

as Gn — e.g., G2, G4, etc.). As shown below, the generalized form is useful in selecting specific spatial-frequency bands to weight for registration, with peak weighting about:

$$f_{peak} = \sqrt{n}/2\pi\sigma_b \tag{3.19}$$

and with frequency band width proportional to $1/\sigma_b$.

3.3 Experimental Methods

3.3.1 Test Images

3.3.1.1 3D-2D: DRR and Radiograph Images

We consider 3D-2D registration (translation-only) of a radiograph to a CT image via DRR. The DRR was generated from an abdominal CT volume (Somatom Definition, Siemens) with a 250 HU soft-tissue threshold and forward projection by trilinear interpolation as illustrated in Fig. 3.1A. Simulated radiographs (Fig. 3.1B) were generated by computing separate forward projections and adding a power-law-distributed random image sample to simulate overlying soft tissue and injecting quantum noise correlated by the system MTF. Two CT noise realizations were generated (as described below in Sec. 3.3.1.2) to ensure that the CT-derived quantum noise was independent between DRRs and simulated radiographs. The method allows generation of many images for performing registration (each having different realizations of soft-tissue content) while maintaining a known ground-truth transformation.

The soft-tissue background was generated from the power-law distribution with $\beta_s = 3.6$ (as in Fig. 3.1B), yielding a distribution that is visually similar and, more importantly,

statistically matches that observed in radiographic images of real anatomy [20]. The background image was scaled and re-centered so that the mean approximated attenuation by 30 cm of water with a standard deviation equal to 5% of the mean. The power-law soft-tissue image was added to the DRR, and quantum noise was simulated using the SPEKTR toolkit [61] to determine the x-ray fluence at the detector for a specified dose, determined by the x-ray tube output (mAs) and beam energy. The transmitted fluence was sampled according to a Poisson distribution to simulate quantum noise, and the image was filtered according to a Lorentzian MTF to simulate scintillator blur [81], yielding simulated radiographs as shown in Fig. 1B. The resulting images were 768×512 pixels with 0.279 mm pixel size. This process was repeated for 100 instances of power-law soft tissue realizations and 11 dose levels (ranging 0.005–500 mAs).

3.3.1.2 3D-3D: Voronoi CT-CT Slice Images

CT slices featuring rigid bone and deformable soft tissue were simulated as illustrated in Fig. 3.2B. Soft tissue was represented by Voronoi distributions from 50 randomly placed seed points in the 512x512 image, each assigned HU values in the range −110 HU to +90 HU in a uniform random distribution. A rigid bone region was inserted using a segmented CT image of a human lumbar vertebra, and the image was cropped to a 32 cm diameter cylinder (typical scale for body CT). To obtain a realization of the same image with soft-tissue deformation, the Voronoi image (prior to inserting the bone segmentation) was subjected to a smooth, random displacement field (Fig. 3.3A) also defined by a low-frequency power-law distribution ($\beta = 4.5$, empirically determined to generate smooth deformations) in displacement vectors in the x and y directions, with α scaled to achieve various magnitudes of deformation.

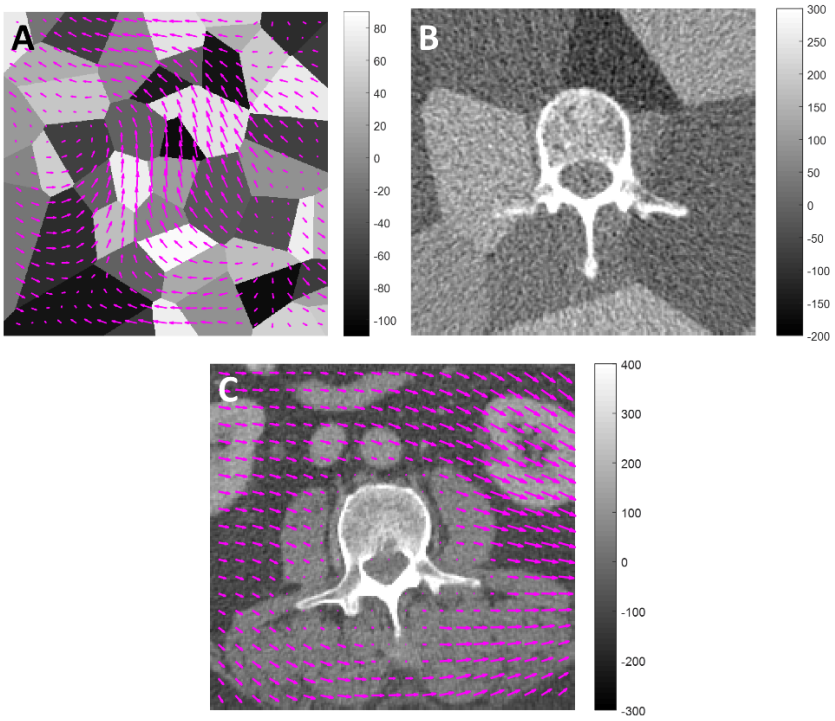


Figure 3.3: Images depicting rigid bone (vertebra) and deformable soft-tissue background. (A) Displacement field overlaid on a Voronoi soft-tissue model. The example shows a mean displacement of 7 pixels (4.7 mm) and interquartile range in displacement 4.4–9.1 pixels (3.0–6.2 mm). (B) Example vertebra + Voronoi image showing a realistic level of correlated noise in CT. (C) Anatomical image (abdominal CT) overlaid with an example deformation field (mean displacement 7 pixels). A mask was applied to ensure rigid motion within the bone region. Figure adapted with permission of the publisher from [77].

Quantum noise in the CT image was simulated by injection of Poisson noise proportional to $1/\sqrt{(1+SPR)\times \text{dose}}$ (with nominal $SPR = 2$). The SPEKTR toolkit [61] was used to determine the fluence for a specified dose for mAs levels ranging 5–1500 mAs (for a 120 kV beam). Projection images (720 images over 360°) were generated from the attenuation values in the CT image and used to compute the expected number of detected photons for each pixel, which was taken as the mean (i.e., lambda) parameter for Poisson sampling at each pixel. Noisy projection images were then reconstructed by filtered backprojection using a Hann apodization filter with a cutoff frequency of $0.8 \times f_{nyq}$. An example image is shown in Fig. 3.3B.

3.3.1.3 3D-3D: Anatomy CT-CT Slice Images

Realistic anatomy depicted in abdominal CT (Fig. 3.2C, a patient image from an IRB-approved study) was regenerated at various dose and deformation levels to test the statistical model on real soft-tissue anatomy. The deformation and noise injection process described in Section 3.3.1.2 was repeated for this CT image, and the region corresponding to the vertebra was masked to ensure zero motion within the bone and smooth reduction of the motion vector field magnitude near the bone boundary (Fig. 3.3C).

3.3.2 Power Spectral Estimates

The power spectrum estimates for signal (G) and noise (S_i, and Q_i) are described below for both 3D-2D and 3D-3D registration scenarios. For given signal- and noise-only images [$g(x,y)$ and $s(x,y)+q(x,y)$, respectively], the power spectrum was estimated by 2D Welch periodogram estimation (3 windows in each direction with 50% overlap) [66] with Hann tapering windows. Models for the power spectra of anatomy [both $G(f_x,f_y)$ and $S(f_x,f_y)$] were based on the estimated periodograms and the well-studied power-law properties of soft tissue described Sec. 3.2.2. Models for quantum NPS were approximations (to reduce the number of parameters for fitting) based on the physical models that describe quantum noise propagation in radiographic [81] and CT [6], [7] imaging systems (more completely described in Section 1.2.1). In radiographic systems, dominant contributors to the NPS are scintillator blur and the detector aperture, the MTF of which may be modeled as a Lorentzian times a sinc [81]. In CT, dominant contributors to the NPS further comprise the ramp filter, apodization filter, and aliasing. The NPS model therefore included a ramp multiplied by the square of the

MTF (Hann apodization filter) and an additive constant. The models and parameters are summarized in Tables 3.1 and 3.2. Parameters for G and S_i were assumed independent of dose, whereas quantum noise parameters were computed at each dose level.

3.3.2.1 DRR (I_1) and Radiograph (I_2)

For the 3D-2D case, the true signal image $g(x, y)$ was given by the DRR, and its estimated periodogram was fit via the model in Table 3.1. The DRR carries a small amount of CT-derived quantum noise which was assumed negligible in fitting $G(f_x, f_y)$ but should still be accounted in $Q_1(f_x, f_y)$. Based on the projection-slice theorem, Q_1 is related to a slice of the CT NPS; however, this CT-derived quantum noise was small in magnitude compared to the signal, and the model simply approximated Q_1 as a constant (c_Q). To determine this constant, two DRRs from two CT instances (Section 3.3.1) were subtracted to yield a noise-only image. A periodogram of the difference image (corrected by a factor of 0.5) was estimated, and the constant was set to be the mean over this periodogram.

The soft-tissue β_S was known from the radiograph simulations (Section 3.3.1.1) leaving the power-law scaling parameter (α_S) and the quantum noise parameters (α_Q) to be fit. Periodograms from 100 radiographs (with DRR subtracted to yield soft-tissue + quantum noise only images) were averaged to obtain power spectrum estimates at each dose level. Fits for α_S and α_Q were performed jointly for the highest dose power spectrum, and the resulting α_S was fixed in fitting α_Q at other dose levels.

Table 3.1: Power Spectrum Models for DRRs and Radiographs

3D-2D: DRR (I_1) to Radiograph(I_2)	
Signal (Bone) Spectrum	$G(f_x, f_y) = \left(\frac{\alpha_G}{f_0^{\beta_G} + f^{\beta_G}}\right) MTF^2$
Soft-Tissue Spectrum	$S_1 = 0, \quad S_2(f_x, f_y) = \left(\frac{\alpha_S}{f_0^{\beta_S} + f^{\beta_S}}\right) MTF^2$
Quantum Noise	$Q_1 = c_Q, \qquad Q_2(f_x, f_y) = \alpha_Q \cdot MTF^2$
MTF	$MTF(f_x, f_y) = \frac{1}{1 + Lf^2} \mathrm{sinc}(f_x, f_y)$

3.3.2.2 CT Slice

The bone-only $g(x, y)$ images for the Voronoi 3D-3D case were formed from the mean of 100 CT images (10 quantum noise realizations for 10 different Voronoi backgrounds) at each dose level. The $g(x, y)$ from the highest dose level was used to compute the periodogram for $G(x, y)$ which was fit to a power-law + exponential function as shown in Table 3.2. This step was repeated to obtain $G(x, y)$ for the bone in the anatomy 3D-3D case using 50 images (each with new noise and deformation).

Based on power spectrum analysis in Sec. 3.2.3.2, $\beta_S = 3$ was used for the soft-tissue parameter value for both the Voronoi and anatomy images. The remaining noise parameters were determined by fits to estimated $S_i + Q_i$ spectra at each dose. Power spectra for each dose level were estimated by averaging the periodograms of 100 CT slices (50 in the anatomical

case) with $g(x,y)$ subtracted (leaving $s(x,y)+q(x,y)$). Again, α_S was determined in a joint fit with the quantum noise parameters at the highest dose level, and the value was fixed when fitting the quantum noise parameters for lower dose levels.

Table 3.2: Power Spectrum Models for CT Slice

3D-3D: CT-to-CT Slice	
Signal (Bone) Spectrum	$G(f_x,f_y)=\left(\frac{\alpha_G}{f_0^{\beta_G}+f^{\beta_G}}+a_G\mathrm{e}^{-b_G f}\right)MTF^2$
Soft-Tissue Spectrum	$S_i(f_x,f_y)=\left(\frac{\alpha_S}{f_0^{\beta_S}+f^{\beta_S}}\right)MTF^2$
Quantum Noise	$Q_i(f_x,f_y)=\alpha_Q\cdot f\ MTF^2+c_Q$
MTF	$MTF(f_x,f_y)=\mathrm{Hann}(f_c)$

3.3.3 Registration Experiments

For each image pair in the following registration scenarios, an initial translation of $\boldsymbol{\tau}=[1.2\text{ pix}, 1.2\text{ pix}]$ was imparted in the moving image prior to registration using cubic B-spline interpolation. (Registration was observed to be insensitive to small changes in initial shift value.) Following the shift, translation-only rigid registration was performed in SimpleITK [63] using each of the similarity metrics (CC, GC, G2, or G4) with σ_b ranging from 1 to 4 pixels in increments of 0.5 pixels. As gradient-based similarity metrics were not implemented in SimpleITK, an analytical equivalent was implemented by noting from Eq. (3.16) that these metrics can be achieved by prefiltering the images to achieve the $H_{Gn}(f_x,f_y;n,\sigma_b)$ frequency weighting in Eq. (3.18) (by filtering both images according to the square root of H_{Gn}) and using

the built-in NCC in SimpleITK. NCC differs slightly from CC in that the images are renormalized at each spatial shift according to the values in the overlapping regions; however, the normalization primarily serves to reduce the influence of local optima rather than improve accuracy at the true solution (which is reflected in the CRLB and RMSE estimate, as both are unaffected by DC shifts and scaling).

For each similarity metric (i.e., n in $H_{Gn}(f_x, f_y; n, \sigma_b)$), the optimal blur was determined by minimizing (2.33) with respect to σ_b, and the observed RMSE at that blur level was compared to both the RMSE predicted by Eq. (2.33) and the CRLB in Eq. (2.12) (which is independent of σ_b). Computation of these figures of merit was achieved using the power spectrum model fits discussed in Sec. 3.2.2. Cases of registration failure were observed to occur for $\sigma_b < 1$ pix or in cases for which f_{peak} [described in Eq. (3.19)] was larger than approximately half the Nyquist frequency; therefore, optimization of σ_b was constrained to satisfy these requirements. Image edge effects introduced by prefiltering were avoided by excluding image boundary regions during similarity metric calculation.

3.3.3.1 3D-2D Registration (effect of dose)

DRR-to-radiograph registration error was examined as a function of radiograph dose, ranging 0.005–500 mAs. For each dose level, 100 simulated radiographs (Sec. 3.3.1.1), each with different quantum and soft-tissue realization, were registered to the bone-only DRR using CC, GC, G2, and G4. RMSE was computed for each dose level and compared to the predicted RMSE and the CRLB.

3.3.3.2 Voronoi 3D-3D Registration (effect of dose)

Voronoi CT-CT slice registration error was examined as a function of dose over the range 5–1500 mAs. For each of 10 Voronoi images, 10 displacement fields (mean displacement magnitude of ~7 pix) were applied to generate 110 CT slices (100 deformed, 10 with original Voronoi) at each dose level. Each of the deformed images was registered to the undeformed slice at the matching dose level using CC, GC, and G4. RMSE at each dose level was compared to the predicted RMSE and CRLB.

3.3.3.3 3D-3D Registration of Anatomical Images (effect of dose)

Registration error of anatomical CT slices was examined as a function of dose over the range 5–1500 mAs. At each dose level, 10 non-deformed noisy images were generated and registered to 10 deformed images generated at the same dose level, yielding 100 registrations for each dose level. The RMSE for CC, GC, and G4 was compared to the predicted RMSE and CRLB. The experiment was performed for two conditions of deformation magnitude with mean displacement magnitude of ~7 pix and ~22 pix.

3.3.3.4 Voronoi 3D-3D Registration (effect of deformation magnitude)

Voronoi CT-CT slice registration error was examined as a function of the soft-tissue deformation magnitude. The experiment of Sec. 3.3.3.2 was repeated (at 250 mAs dose level) for 12 levels of displacement field magnitude by varying α in the power-law derived displacement fields to yield mean pixel displacement magnitude ranging from ~0.01 to 22 pix.

Registration results were compared to RMSE predictions and RMSE measurements in registered images containing different Voronoi backgrounds (such that the soft-tissue noise terms were truly independent) to check the extent of deformation necessary to justify the assumption of independence. Registrations were performed for each of the 10 no-deformation CT slices (each with a different Voronoi background), yielding 45 (i.e., 10-choose-2) registrations to examine RMSE for CC, GC, and G4.

3.3.3.5 3D-3D Registration of Anatomical Images (effect of deformation magnitude)

Registration error of anatomical CT slices was examined as a function of the soft-tissue deformation magnitude. The experiment Sec. 3.3.3.3 was repeated (at the 250 mAs dose level) for 14 levels of displacement field magnitude by varying α in the power-law displacement fields to yield mean pixel displacement magnitude ranging from ~0.01 to 22 pix. Registration results were compared to RMSE predictions for each similarity metric (CC, GC, and G4).

3.3.4 Model Exploration: Effect of Soft-Tissue Parameters (α_S and β_S)

The soft-tissue power-law parameters (α_S and β_S) can vary as a function of the contrast and texture, respectively, of the soft-tissue. Increasing α_S leads to a greater soft-tissue intensity range. Increasing β_S leads to cloudier, more smoothly varying texture, whereas reducing β_S yields higher-frequency content (and $\beta_S = 0$ giving white noise). Such texture changes may be associated with changes in anatomical region (e.g., abdominal anatomy vs. lung parenchyma) or the use of a different imaging modality (e.g., radiography vs. CT or

ultrasound). To understand the role of these parameters on registration performance, we used Eq. (2.33) to predict the registration error for CC, GC, and G4 (at optimal σ_b) as a function of these soft-tissue power-law parameters. For both 3D-2D and 3D-3D scenarios, we fixed the model parameters described by Tables 3.1 and 3.2 at several dose levels and separately varied α_S and β_S. As α_S values are not comparable for different values of β_S, the α_S value was scaled to achieve the same area under the curve (energy) of the original power-law distribution.

3.4 Results

3.4.1 Registration Results: Comparison of Theory and Measurement

Figure 3.4A shows 3D-2D registration error as a function of dose for four similarity metrics. Solid lines depict the predicted RMSE via Eq. (2.33) for each metric at optimal σ_b (computed for each dose level and metric), and the markers represent the experimental error for that value of σ_b. Immediately apparent is the large performance gap between CC and gradient-based metrics, with CC showing more than an order-of-magnitude greater error than the other metrics. Further, CC performance appears to be soft-tissue-limited in that increased dose (and thus reduced quantum noise) does not yield improved registration accuracy. For the gradient-based metrics, however, RMSE decreases as a function of dose over the range ~0.005–1 mAs and follows the trend set by the CRLB (dashed line). For higher dose, a plateau in RMSE is exhibited for all metrics (and the CRLB), again indicating that the registration is limited by soft-tissue noise rather than quantum noise. The best registration error was obtained using the G4 metric, giving RMSE = 0.006 pix, (compared to the CRLB = 0.003 pix) at the 500 mAs dose level.

Fig. 3.4B shows Voronoi 3D-3D registration error as a function of dose for CC, GC, and G4 similarity metrics in the presence of soft-tissue deformation. Each metric exhibits a similar plateau as seen above; however, CC plateaus at a much lower dose level than GC, G4 (~25 mAs vs. ~1000 mAs). Interestingly, GC (only slightly outperforming G4) nearly achieves the CRLB over all dose ranges tested, indicating near optimality as a metric for the soft-tissue deformation scenario.

Figures 3.4C and 3.4D shows the error in 3D-3D registration of anatomy in the presence of soft-tissue deformation with mean deformation magnitude of 7 mm and 22 mm, respectively. RMSE is shown as a function of dose for CC, GC, and G4 similarity metrics. Interestingly, the agreement between theory and measurement improves with the magnitude of displacement — with predictions underestimating the measurements at 7 pix displacement and agreeing well for larger displacement (e.g., 22 pix deformation). It is important to note that the predicted RMSE is identical for the two plots in Figs. 3.4C and 3.4D, showing that the measured RMSE for CC and GC improves greatly in the presence of *increased* soft-tissue deformation. Meanwhile, the G4 metric shows good agreement between measurement and prediction for both the small and large deformation scenarios. Together, this indicates that small deformations in the case of real anatomy (cf., the sharp edge scenario of the Voronoi images) do not sufficiently decorrelate the soft-tissue background for the CC and GC metrics, and the lack of non-correspondence in soft-tissue backgrounds degrades the search space. The G4 metric, however, emphasizes finer gradient features, and a smaller magnitude of deformation is sufficient for corresponding background structures to become uncorrelated, thereby improving the search space quality and improving registration accuracy.

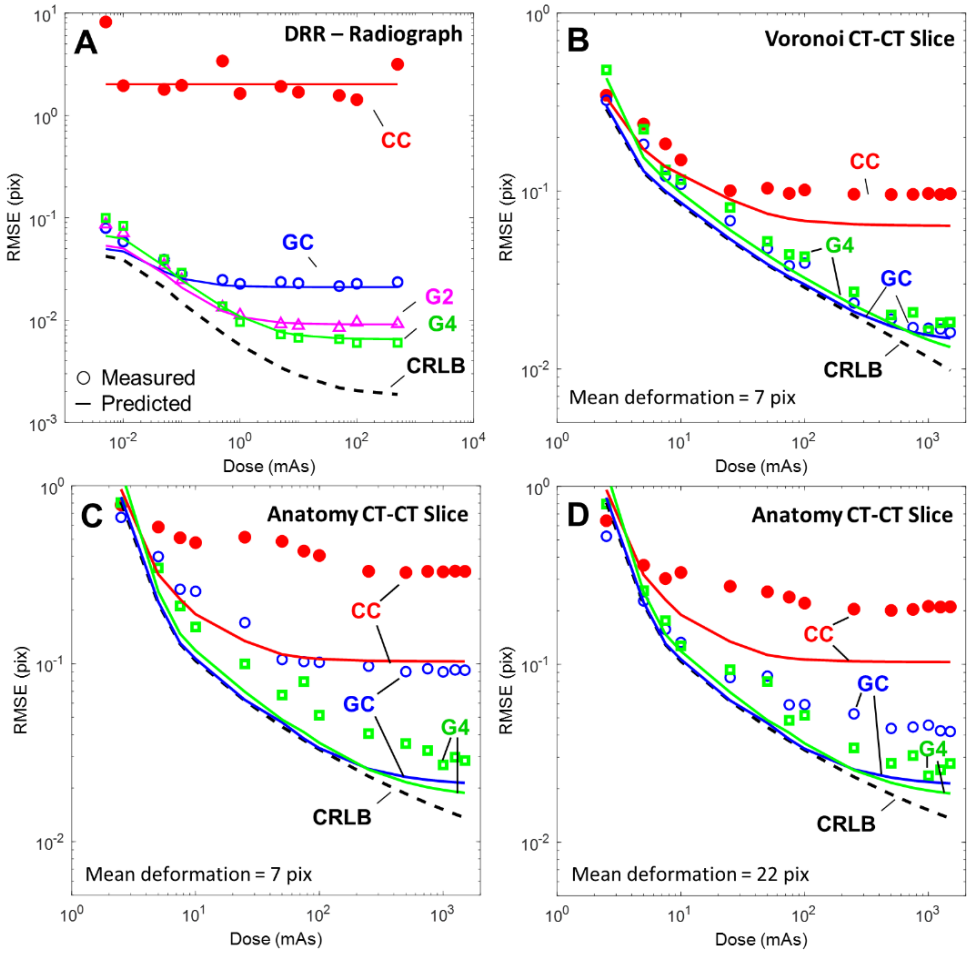


Figure 3.4: Effect of dose on registration performance for (A) 3D-2D registration and (B) Voronoi 3D-3D registration with 7 pix mean deformation, and Anatomy 3D-3D registration with (C) 7 pix mean deformation and (D) 22 pix mean deformation. Each plot shows the predicted error for each metric at optimal σ_b (solid lines), the measured error for each metric at that σ_b (markers), and the CRLB (dashed line). Similarity metrics examined included CC (red), GC (blue), G2 (magenta), and G4 (green). Figure adapted with permission of the publisher from [77].

It is important to keep in mind that for both scenarios in which soft tissue presents as a source of "noise" (i.e., soft-tissue absence in 3D-2D and soft-tissue deformation in 3D-3D), the model predictions were achieved by simply incorporating soft-tissue as a power-law noise distribution in Eqs. (2.12) and (2.33). Further, in both scenarios the predictions and experiments showed improved performance when using the gradient-based similarity metrics. This is particularly interesting when compared to results in the following section which show that CC outperforms GC when no deformation is present. To understand this change in the

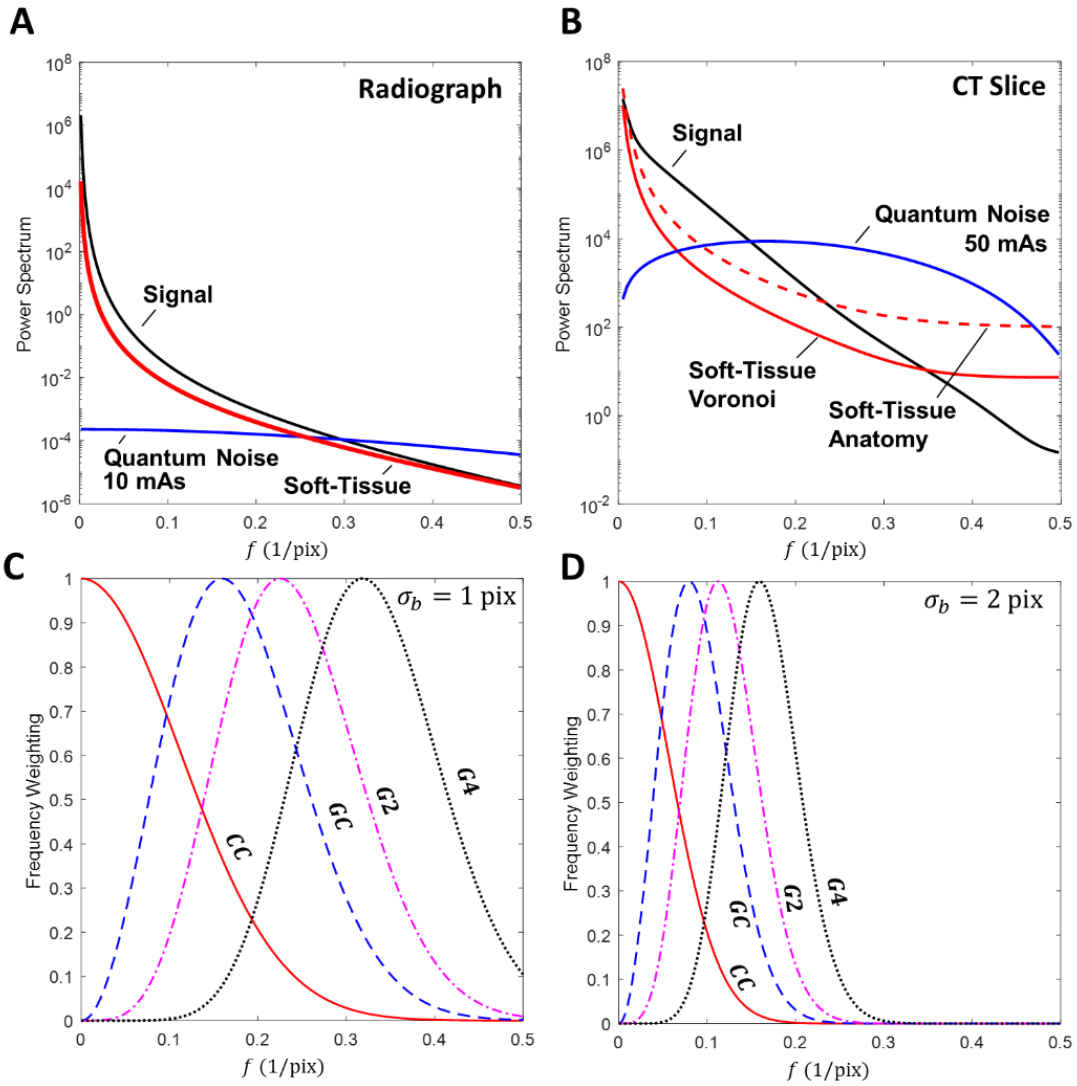


Figure 3.5: Power-spectrum profiles for the signal (black), soft-tissue (red), and quantum noise (blue) terms fit to (A) Radiograph (10 mAs) and (B) Voronoi CT slice (50 mAs) image data (with an additional dashed line profile of the soft tissue anatomy spectrum) using the models in Tables 3.1 and 3.2. Registration frequency weighting profiles using Eq. 10 for CC (red), GC (blue), G2 (magenta), and G4 (black) at (C) $\sigma_b = 1$ pix and (D) $\sigma_b = 2$ pix. Figure adapted with permission of the publisher from [77].

preferred metric, it is important to examine the power spectra of both the signal and noise terms as seen in Fig. 3.5 and to compare these spectra with the frequency weighting that each metric provides. In the presence of quantum noise alone, it is clear from Fig. 3.5A–B that there is a large signal-to-noise ratio near the zero-frequency region; therefore, it is intuitive that CC (which weights the low-frequency band) is the preferred metric. However, in the presence of soft-tissue deformation (modeled as low-frequency noise), the low-pass nature of the power-

law soft-tissue spectrum leads to a sharp reduction in signal-to-noise ratio near zero frequency. Therefore, the use of gradient-based metrics, which down-weight the near-zero frequency regions, is preferred.

3.4.2 Effect of Deformation Magnitude

Fig. 3.6A shows Voronoi CT-CT slice registration error as a function of the mean magnitude of pixel displacement in deforming soft tissue. The results are compared with the dashed lines that show predicted RMSE and dotted lines showing experimental registration performance for images with different realizations of Voronoi background (i.e., independent soft-tissue noise terms). The lowest registration error was observed for cases of minimal deformation, since soft-tissue anatomy contributes to accurate alignment in such cases. Furthermore, in the absence of deformation (in which case the underlying images differ only by quantum noise), CC is found to be the optimal metric. However, as deformation magnitude increases, registration error increases up to a plateau near ~5–6 mean pixel displacement, showing that beyond a certain level of deformation, when the soft-tissue backgrounds are sufficiently decorrelated, the magnitude of deformation has little effect on the registration error. For both metrics, the plateau occurs at the error level observed when registering newly-generated independent Voronoi backgrounds, supporting the assumption that, under large deformations, the soft tissue can be treated as an independent noise term. It is also interesting to note the hump in RMSE for GC, which is attributable to local optima created when gradient-based metrics are used in the presence of small deformation.

Figure 3.6B similarly examines the impact of deformation magnitude for CT-CT slice registration of real anatomy. A similar behavior to the Voronoi results is observed for deformation <5 pix, and the measured RMSE values plateau at much higher levels of deformation than in Fig. 3.6A, particularly for the CC and GC metrics. Interestingly, it appears that the speed of convergence is related to the metric order, with G4 plateauing faster than GC, which in turn converges faster than CC. We observe this effect also in Figs. 3.4C–D, where smaller deformation magnitude was needed for the soft-tissue background to be sufficiently decorrelated when using higher-order gradient metrics.

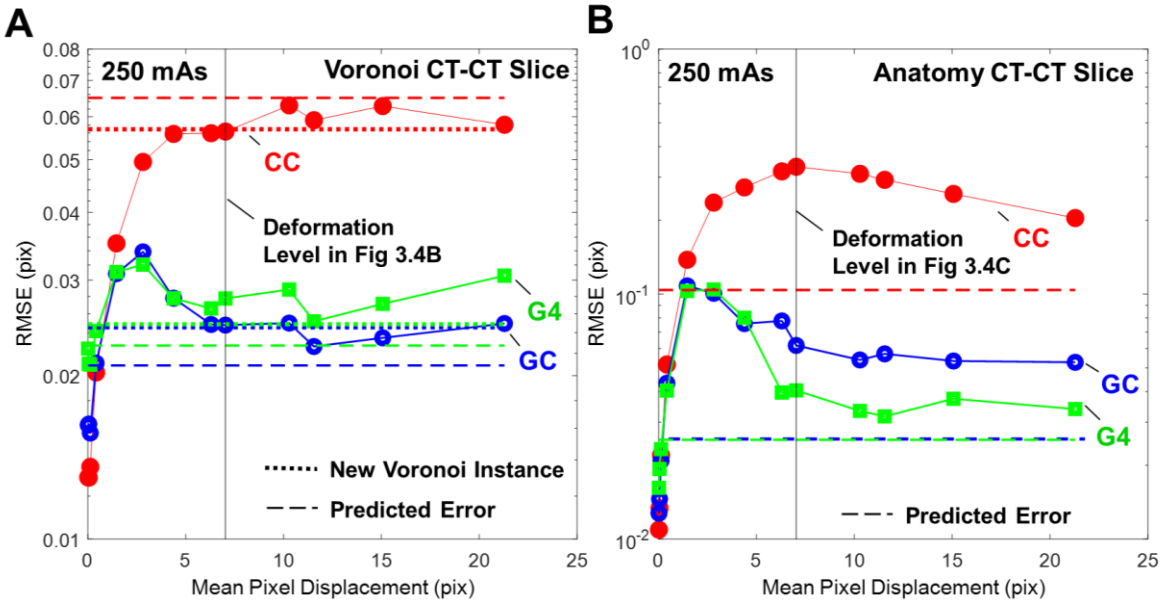


Figure 3.6: 3D-3D registration error as a function soft-tissue deformation magnitude for CC (red, solid circle), GC (blue, open circle), and G4 (green square) for (A) Voronoi and (B) anatomy CT-CT slice registration. Dashed lines show the predicted registration performance of Eq. (2.33) for each metric. Dotted lines in (A) depict the registration performance for each metric when registering CT slices that contain different (independent) instances of Voronoi soft-tissue background. Figure adapted with permission of the publisher from [77].

3.4.3 Effect of Soft-Tissue Parameters (α_S and β_S)

Figure 3.7A shows the predicted 3D-2D RMSE at optimal σ_b for CC (red), GC (blue), and G4 (green) as a function of soft-tissue contrast magnitude (α_S) for 2 dose levels. At small α_S (thus dominated by quantum noise) CC slightly outperforms the others and registration performance is quantum limited, with changes in α_S having little or no effect on registration. As α_S increases, however, (yielding stronger contrast from soft-tissue) GC becomes the preferred metric due to its effective down-weighting of low frequency noise content. As α_S becomes large, G4 becomes the preferred metric and the RMSE converges for all dose levels, indicating that the performance is limited by soft-tissue deformation. Similar behavior is observed in Fig. 3.7B for 3D-3D registration.

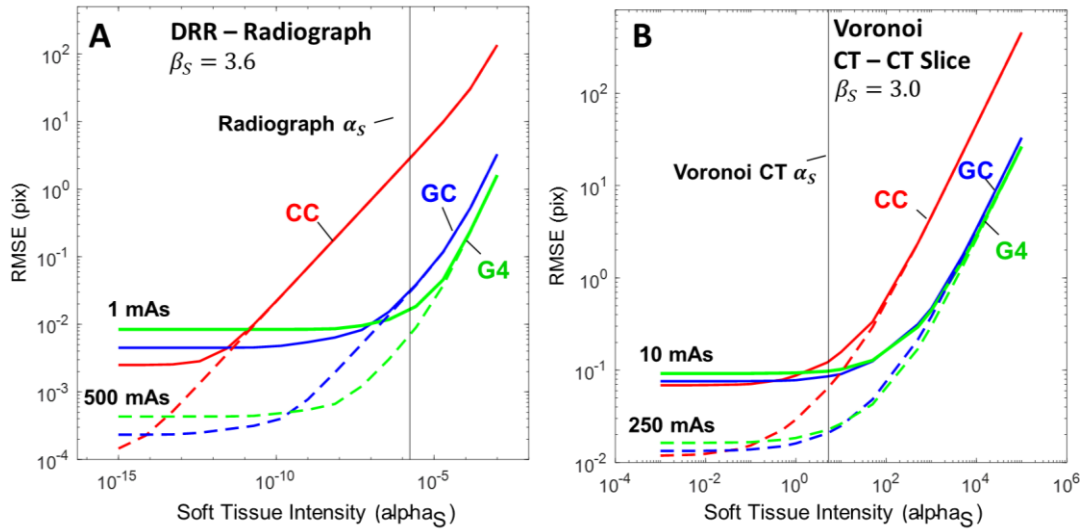


Figure 3.7: The effect of the deformed soft-tissue contrast term, α_S, on registration performance. Predicted RMSE at optimal σ_b shown for CC (red), GC (blue), and G4 (green) at various dose levels for (A) DRR-Radiograph and (B) Voronoi CT-CT slice registration. Figure adapted with permission of the publisher from [77].

Figure 3.8A shows the effect of β_S (at fixed total power) on the performance of CC, GC, and G4 similarity metrics for 3D-2D registration. At $\beta_S = 0$ (i.e., white noise) CC is the preferred metric, since the NPS does not peak near zero frequency. As β_S increases, however, soft-tissue noise occupies the same frequency region as the signal term, leading to increased error for all metrics. For further increase in β_S (and with the soft-tissue power spectrum concentrated near zero frequency) we see that error decreases for the GC and G4 metrics, since they effectively attenuate the lower-frequency soft-tissue noise. The performance of CC, however, plateaus at a much higher RMSE and has no dose dependence, illustrating that soft-tissue deformation dominates CC registration performance. Figure 3.8B shows a similar non-monotonic trend for the GC and G4 case in 3D-3D registration. Interestingly in both scenarios, the highest CRLB error is seen for β_S in the range of 1–2. This can be understood by comparison of Eq. (2.12) with signal power spectra of Fig. 3.5A–B, where we see from the f_j^2 term in Eq. (2.12) that higher frequencies provide (quadratically) more information in registration, whereas the DC component provides no information. However, in Fig. 3.5A–B we see that the signal power spectrum is concentrated in the low frequency range, which combined with the f_i^2 weighting, implies that the mid-to-low frequencies effectively provide the most information for registration. Therefore, soft-tissue noise with β_S ~1−2 presents the most confounding influence in the mid-to-low frequency range. Higher values of β_S concentrate the noise in the low-frequency region, and lower values of β_S pushes the noise to the higher frequency range, where there is little signal power.

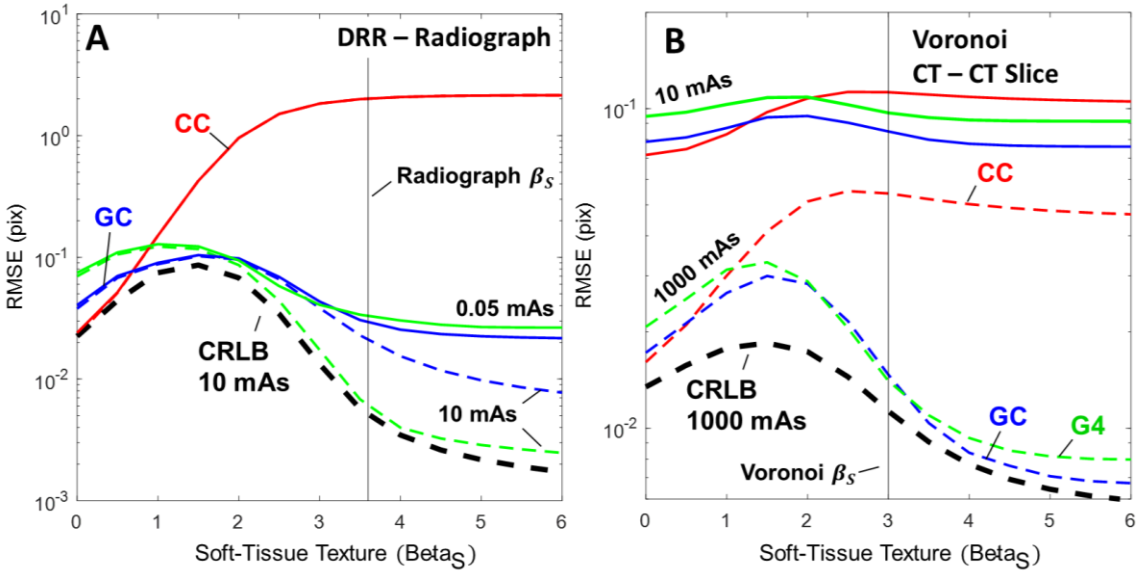


Figure 3.8: The effect of the deformed soft-tissue texture term, β_S, on registration performance. Predicted RMSE at optimal σ_b shown for CC (red), GC (blue), and G4 (green) at various dose levels for (A) DRR-Radiograph and (B) Voronoi CT-CT slice registration. Figure adapted with permission of the publisher from [77].

3.5 Conclusion

In this chapter, a model for rigid registration performance was presented in which soft-tissue deformation was incorporated as a noise source. By adopting concepts from signal detection theory in modeling soft tissue as a noise-power spectrum with power-law spatial frequency dependence and incorporating it in a statistical framework for registration error, the influence of factors such as dose, noise, and choice of similarity metric on registration performance were quantified. In particular, CC-based and gradient-based metrics were shown to differ according to their frequency domain weighting of the signal, quantum noise, and soft-tissue power spectra, where gradient-based methods were shown to best compensate for the presence of soft-tissue deformation.

We further investigated the magnitude of soft-tissue deformation consistent with the assumption of independent noise in the model for rigid registration performance. Of course, soft-tissue deformation is not a random process, but the abstraction was shown to hold

reasonably well for large deformations giving rise to large regions of non-corresponding tissue overlap. This in turn was shown to be modeled well as noise for various similarity metrics. We further showed (Fig. 3.6A) that large deformations yield the same RMSE as registration of images with independent realizations of soft-tissue background (where the independent noise source assumption is clearly valid), supporting the notion that large soft-tissue deformation may be considered as independent noise in rigid registration. It is important to note that deformation magnitude should be much larger than the correlation length of the soft-tissue gradient image (such that high-gradient regions are no longer overlapping). The study shown above (Sec. 3.4.2) investigated the magnitude of deformation required to justify this claim for Voronoi images in the 3D-3D case, where mean deformation magnitudes > 5 pix yielded the same error as the independent background case. However, the Voronoi images contain sharp gradients which have small correlation length (on the level of system blur, ~2 pix) due to the step-function nature of the Voronoi model. For the case of CT-CT registration of real anatomy case (which exhibited somewhat longer-range correlations in the gradient images compared to the Voronoi case) larger deformations were necessary to support the assumption of independence. Interestingly, however, the long-range gradient correlations in such images were suppressed by gradient-based similarity metrics (especially G4), thereby greatly reducing the magnitude of deformation that was necessary for the independence assumption. Finally, it is important to note that the independence assumption is not necessary in the 3D-2D registration case, since soft tissue is only present in one of the images.

The method for simulating soft-tissue deformation in this work (Sec. 3.3.1) involved a random displacement that was not physically or biomechanically motivated and may imply somewhat unrealistic deformation characteristics. For example, since the displacement fields

were randomly generated from a power-law distribution, there is no guarantee that the transformations are diffeomorphic; despite this, we observed that the method did indeed exhibit diffeomorphic properties (positive Jacobian determinant) over the range of deformation magnitude considered. (Non-diffeomorphic fields were observed for mean pixel displacement greater than 23 pix). Another potential limitation in the simulation is the lack of a biomechanical model to constrain deformation magnitude — for example, constraining deformation to be small near bone-tissue interfaces (attachment). Doing so would suggest that some soft-tissue (i.e., that near bone) should not be treated as noise and should be included as salient features for registration. A simple method to accomplish this would be to split the soft-tissue power spectrum across N and G, with $N_i(f_x, f_y) = Q_i(f_x, f_y) + (1 - a)S_i(f_x, f_y)$ and $G(f_x, f_y) \leftarrow G(f_x, f_y) + aS_i(f_x, f_y)$, where $a \in [0,1]$ represents the portion of non-deformed soft tissue. However, such a model is outside of the scope of this thesis.

The equations in Tables 3.1 and 3.2 represent anatomy, soft-tissue clutter, and quantum noise described by circularly symmetric power spectra for purposes of simplicity. The isotropic assumption is not central to the methods described above, and while such models provided reasonable fits to the experiments conducted in this work, anisotropic power spectra can certainly be incorporated in the framework. Scenarios that may warrant such models include anatomy presenting strong directionality (e.g., 3D ductal breast tissue [82]) or CT quantum noise that can be strongly correlated in non-circular objects and/or with x-ray tube mA modulation techniques.

In this chapter, the statistical framework describes the translation-only case in order to gain basic insight into more general scenarios. While the effect of soft-tissue deformation on rigid registration was examined in this work, it is important to note that the current framework

does not apply to deformable registration. In Chapter 5 we will discuss how the framework may be extended to scenarios of deformable registration in which both bone and soft tissue present salient information in the registration process.

The experiments in this work examined x-ray projection (3D-2D) and CT images (3D-3D). The framework, however, is certainly generalizable in the 3D-3D case to other same-modality registration scenarios (e.g., magnetic resonance, MR-MR, or ultrasound, US-US), as long as the underlying image content is consistent, and the noise is properly characterized. While the projection-based 3D-2D registration is somewhat unique to the scenario of radiographs and CT, the model may generalize to other 3D-2D scenarios (e.g, US slice-to-volume). In the following chapter, we delve deeper into the topic of 3D-2D (DRR-Radiograph) registration to show how the insights of this model are reflected in application to a registration method for automatically labeling vertebrae in a radiograph.

Chapter 4: Deformable 3D-2D Registration for Image-Guided Spine Surgery

4.1 Introduction

In the previous chapter, a statistical model was developed that provided an important basis for understanding how anatomical deformation and/or mismatch in image content acts as a noise source in rigid image registration. The model also provided guidance for how to compensate or mitigate these factors. Most notably, gradient-based similarity metrics were shown to provide clear advantage over intensity-based metrics for robust registration performance in the presence of deformation. In this chapter, the implications of the model are demonstrated in application to 3D-2D image registration to accurately label vertebrae in radiographic images even in the presence of deformation. The registration problem involves challenges of quantum noise and anatomical deformation, reflecting the insights gained from the theoretical model derived in Chapters 2 and 3.

In image-guided spine surgery, target localization using 2D intraoperative radiographs is an essential step in effective treatment. However, accurate interpretation of anatomy in radiographs during surgery can be a challenging, time-consuming, and error-prone task that can confound even experienced surgeons. In the case of vertebral level identification, surgeons typically acquire multiple radiographic images at shifted fields of view to "count" to the correct level — a process involving considerable time and stress in the operating room to ensure accurate localization. Some institutions may implement an additional preoperative procedure for tagging (injection) of radio-opaque cement into the target vertebra under CT guidance to allow fast, unequivocal identification in the operating room. Even so, wrong-level spine surgery is reported to occur in approximately 1 in 3110 spine surgeries (and up to 1 in 700 for lumbar disc procedures), and it is estimated that up to half of spine surgeons will encounter this error at some point in their career. A wrong-level error can lead to suboptimal (or failed) surgical product and potentially costly litigation [83]. To prevent such occurrences and potentially improve workflow and confidence in target localization, a 3D-2D registration framework (called LevelCheck) has been shown to automatically overlay relevant target anatomy such as vertebral labels from 3D preoperative imaging (CT or MR) to the intraoperative 2D image as a means of decision support [79], [80], [84], [85].

LevelCheck operates as a rigid 3D-2D registration method that estimates the rigid 6DOF transformation (T_r) of the CT volume within the virtual source-detector geometry depicted in Fig. 4.1A. Following estimation of T_r, vertebral labels defined in the CT volume are projected onto the radiograph as in Fig. 4.1B. The method faces many challenges relevant to the statistical model presented in Chapter 3, including poor image quality (quantum noise) and content mismatch — where both soft tissue and surgical instrumentation are present in the

radiograph but not the DRR. In line with our theoretical analysis, gradient-based similarity metrics were shown to be robust to these effects [80], allowing accurate rigid registration as illustrated in the example of Fig. 4.1B; however, the accuracy of rigid 3D-2D registration may be compromised due to the presence of deformation.

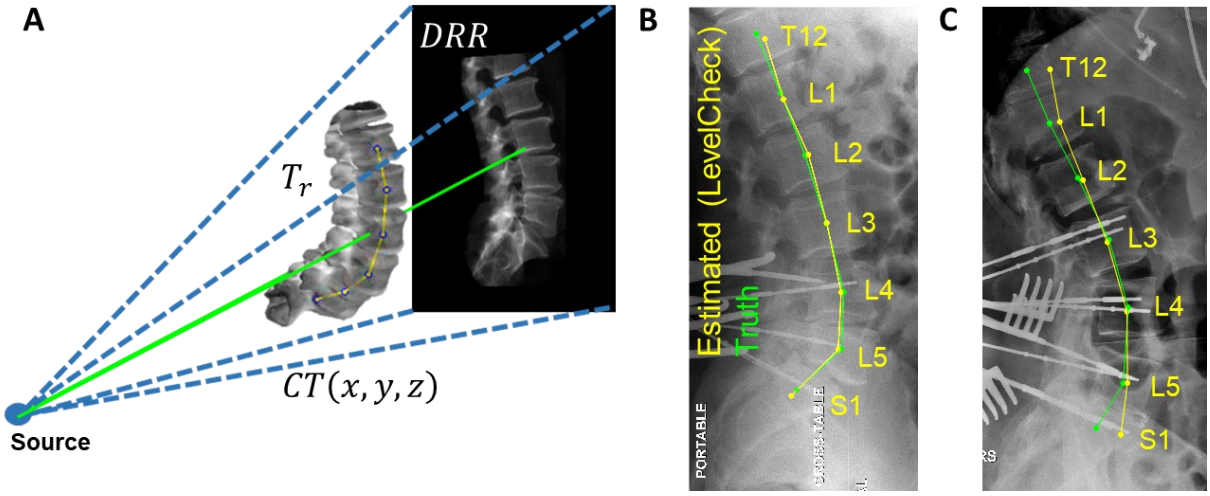


Figure 4.1: (A) 3D-2D projection geometry by which a DRR is generated from a preoperative CT oriented according to the 6DOF pose, T_r. (B–C) Example LevelCheck registrations (yellow) compared to radiologist-defined true positions of the vertebrae (green). (B) Case showing good registration according to a rigid model. (C) Case with a strong change in spinal curvature for which the conventional rigid approach shows a degradation in registration accuracy at the superior and inferior extent of the radiograph. Figure adapted with permission of the publisher from [86].

Spine deformation is particularly common due to the differences in preoperative and intraoperative patient positioning, where 3D preoperative images are typically acquired with the patient lying supine on the scanning bed, whereas intraoperative images are often acquired with the patient prone on an arched surgical table (e.g., Jackson table or Wilson frame). The effect of such deformation on rigid 3D-2D registration is illustrated in Fig. 4.1C, showing accurate alignment in the central region of the image that diminishes in the superior and inferior regions due to changes in spinal curvature between the preoperative and intraoperative images. Recent clinical studies [85] found a reduction in surgical confidence arising even from fairly

small errors for which the label was still within cortical boundaries but not near the vertebral centroid; therefore, a method to address and compensate for this deformation is necessary.

While 3D-3D deformable registration is a topic of widespread research, the topic of deformable 3D-2D image registration presents a challenge that remains to be fully addressed. Previous work has illustrated that deformable 3D-2D registration is a challenging and often ill-posed problem; for example, in the case of a single 2D view there is degenerate geometric relationship between projection magnification and changes in object size. Several advances in deformable 3D-2D image registration have been reported for scenarios in which multiple 2D images are acquired [87]–[91], in applications such as radiographic vertebrae segmentation and surgical guidance using digital subtraction angiography. For single-view 3D-2D registration, at least two general methods to mitigate the effects of deformation have been proposed: (i) apply a deformable method with constraints imposed by 3D segmentations and deformation models [92], [93]; and (ii) perform piece-wise rigid registrations, often involving shape models [94] or segmentations of rigid bodies. In the context of spine registration, the segmentation methods often utilize 3D vertebral segmentations and perform registration for each vertebra individually [95], [96].

To provide improved accuracy under conditions of spinal deformation, we propose a method that is analogous to a combination of two common forms of image registration: (i) piecewise rigid registration (as posed in Penney [95], for example), although our method does not rely on an explicit segmentation of "pieces"; and (ii) block matching (as in Ourselin *et al.* [97] and Zhu and Ma [98], for example), where our block definitions are based on relevant anatomical structures using the label annotations specified in the preoperative 3D image, rather than arbitrarily dividing the image into sub-images. Furthermore, our method reflects and

extends that of Varnavas *et al.* [99], in using information from multiple single vertebra registrations; however, we define sub-image masks that incorporate subsets of adjacent vertebrae, rather than single vertebra masking. Moreover, the method is automatic (the same level of automation as with LevelCheck), and beyond the definition of vertebral labels in the preoperative image (which can be done manually or automatically [100]), it requires only an initialization in the superior-inferior direction of the CT.

In this chapter we present a single-view, multi-stage, 3D-2D registration method that robustly accounts for spinal deformation by rigidly registering sub-images of decreasing size at each subsequent stage. The multi-stage method is referred to as msLevelCheck. The method yields a 3D-2D registration that is locally rigid (with respect to the registration of any particular sub-image and label annotation therein) yet globally deformable (with respect to the overall motion of all label annotations) and results in accurate target localization over the entire field of view. The method is detailed below, focusing on two main aspects of the algorithm: the generation of sub-image masks from the (pre-existing, automatically computed) annotation locations in 3D without segmentation of the vertebrae; and the multi-scale framework for registration of increasingly local rigid regions. The method is evaluated in phantom experiments as well as a clinical study of patients undergoing thoracolumbar spine surgery.

The work appearing in this chapter was reported in the following journal paper: (M.D. Ketcha et al., *Phys. Med. Biol.,* 62(11), 2017) [86].

4.2 Methods

4.2.1 Rigid 3D-2D Registration Framework

As with the conventional, rigid LevelCheck method [79], [84], msLevelCheck intends to aid target localization by mapping vertebral labels defined in the preoperative CT image (or MR image [101]) to the intraoperative radiograph via image-based 3D-2D registration. Rigid registration is performed by determining a rigid 6DOF transformation of the CT image that optimizes the similarity between the DRR and the intraoperative radiograph (p). The resulting transformation enables the vertebral labels in the CT image to be accurately projected and overlaid in p. The basic LevelCheck framework is illustrated in Fig. 4.2, consistent with Otake *et al.* [79], [84] and De Silva *et al.* [80].

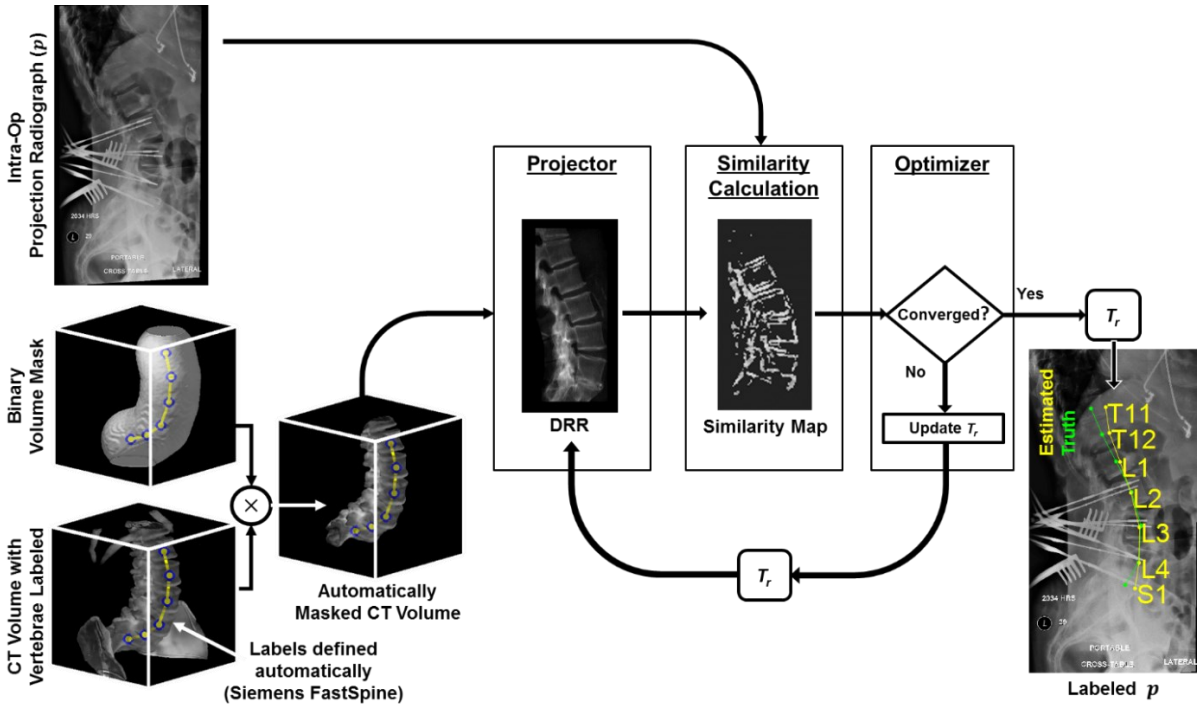


Figure 4.2: Flowchart for LevelCheck 3D-2D rigid registration. Figure adapted with permission of the publisher from [86].

4.2.1.1 Binary Volume Masking

Mismatch of anatomy about the spine, such as the ribs, pelvis, and skin line, can challenge robust 3D-2D registration. To mitigate the effect of such surrounding anatomy and to focus the registration on the spine, we introduced a method for automatically masking the CT volume using already defined 3D vertebral label positions as shown in Fig. 4.2. In this approach, a 3D linear interpolation of the label positions is computed, and a binary volume mask (scale factor of 0 or 1) is defined to include voxels within a distance r of the interpolated line. Previous work [102] identified a nominal distance of r = 50 mm, which was used in the studies reported below. Furthermore, this masking is combined with a soft tissue threshold of 150 HU (setting the value to 0 if below) to focus the registration on the bony vertebral anatomy.

4.2.1.2 Projection geometry and DRR formation

Forward projections were computed within a fixed camera geometry with a virtual detector centered at the origin and an x-ray point source positioned at (x_s, y_s, z_s) with z_s defined to be perpendicular to the detector plane, as described in [84]. A geometry with the piercing point at the center of the detector and a source-to-detector distance (SDD) of 100 cm was assumed. A rigid 6DOF transformation, T_r, consisting of 3 translations (x_r, y_r, z_r) and 3 rotations $(\eta_r, \theta_r, \phi_r)$ defined the pose of the CT volume within this camera geometry. Given the geometry and CT position, a projective transformation matrix, $T_{3\times4}$, was defined to map a

location in the CT coordinate system (x, y, z) to its projected location in the DRR (u, v) according to:

$$c\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = T_{3\times4}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} z_s & 0 & x_s & 0 \\ 0 & z_s & y_s & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} & R_{3\times3}(\eta_r, \theta_r, \phi_r) & & \begin{matrix} x_r - x_s \\ y_r - y_s \\ z_r - z_s \end{matrix} \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (4.1)$$

where c is a constant that normalizes the third element of the 2D position vector. The DRR was generated via ray-tracing [103] with line integrals computed using trilinear interpolation. To achieve pixel-wise correspondence, the virtual detector was defined to have dimensions and pixel size identical to the projection image, which has been resampled to a specified isotropic pixel size (a_{pix}) and rectangularly cropped to exclude collimator edges and burnt-in text annotations. Assuming a nominal magnification factor of 2, the volume was downsampled isotropically to $a_{pix}/2$. The step length for ray casting was chosen to be 2 voxels (equivalently, a_{pix}) based on a sensitivity study reported in Otake *et al.* [79]. Basic overlap of the DRR with the projection image was ensured by translating the CT volume along the longitudinal direction of the patient, thereby determining an initial value for y_r, and resulting in an initial registration error of ~20–200 mm. To improve computation time, DRRs were computed using a parallelized implementation in CUDA on GPU (nVidia, Santa Clara, CA).

4.2.1.3 Similarity Metric

Content mismatch stemming from soft tissue and surgical instrumentation prompts a careful consideration when choosing the similarity metric. From the previous chapter, we know

that gradient-based metrics offer strong advantages with respect to these factors; however, experimental evidence [80] has shown GC in particular to be susceptible to local optima caused by the sharp metal gradients. Therefore, based on the work of De Silva *et al.* [80], we utilize Gradient Orientation (GO), which was shown in to provide a high degree of robustness against this content mismatch by minimizing the impact of signal intensity on the similarity metric calculation and instead focusing on the alignment in gradient direction. The GO similarity was defined as:

$$GO = \frac{1}{\max\left(N, N_{LB}\right)} \sum_{\{i:\, \nabla DRR_i > t \,\cap\, \nabla p_i > t\}} w'(i)$$

$$\text{where} \quad w'(i) = \frac{2 - \ln(|\theta_i| + 1)}{2} \tag{4.2}$$

and reflects the pixel-wise similarity in gradient direction, w', among pixels whose gradient magnitude passes a threshold t in both images, defined as the median gradient intensity. Here, θ_i is the angle difference (radians) in gradient direction between the DRR and p at pixel i. The normalization constant N is the number of pixel locations for which the metric is evaluated, and N_{LB} is a lower bound cutoff set to 30% of the total number of pixels to penalize low counts associated with poor overlap (shown in previous studies to be a stable, nominal parameter setting for this application in [80]).

4.2.1.4 Optimization

The GO metric was optimized over the six-dimensional search space to find the transformation T_r. Due to the highly non-convex nature of the objective space, a multi-start

CMA-ES [104] optimization was employed (Section 1.3.3). Assuming poor initialization, we incorporated multi-starts in which parallel optimizations were performed with initializations distributed over the entire optimization search range. To distribute these initializations, a plane-splitting *kD* tree partitioning of the search space [105] was implemented where the search space was divided by iteratively splitting the largest current subspace in half. The number of multi-starts ($MS = 50$), the population sampling size ($\lambda = 125$), and the search range (*SR*) along each 6DOF dimension ($\pm$ [100 mm, 200 mm, 75 mm, 15^o, 10^o, 10^o]) were selected based on a sensitivity study using a clinical image dataset, considering trade-offs in computation time, robustness, and initialization error. The chosen *SR* values reflect the assumption of a coarse estimate longitudinal initialization but fairly accurate rotational initialization that comes from knowledge of patient positioning (e.g., knowing that the image is a lateral radiograph). As detailed in [79], [84], a rigorous parameter sweep was performed to determine values for *MS* and lambda that yielded robust performance while minimizing graphics processing unit (GPU) memory and computation time.

Convergence was defined with respect to the covariance matrix, where tolerance cut-off (TolX) was set for the maximum value of the diagonal terms of the covariance matrix, implying that convergence is met when population sampling is contained to a small subspace around the current estimate. Due to the wide distribution of the multi-start initializations, it is expected that a large portion of the optimizations may converge to a false optimum; thus, it is computationally inefficient to set a strong convergence criteria. Therefore, a weak tolerance (TolX = 1) was initially set for all the multi-start optimizations, and following convergence for each of these *MS* optimizations, a single-start optimization restart was performed using the highest GO solution as an initialization and a stronger convergence criteria (TolX = 0.1).

4.2.2 Multi-stage LevelCheck framework

From Chapter 3, we know that anatomical deformation not only inhibits accurate rigid alignment — leaving some labels misplaced in the area of deformation — but also reduces registration accuracy in regions that are properly registered (e.g., the L3 and L4 in Fig. 4.1C). Interestingly, in this case the bone anatomy itself is globally deformed due to the articulated nature of the vertebrae; therefore, while some vertebrae are properly registered, the remaining misaligned vertebrae (e.g., T12–L2 in Fig. 4.1C) act only to confound the registration. Therefore, to account for anatomical deformation between the CT and radiographic image, we developed a multi-stage registration framework, henceforth referred to as msLevelCheck.

The core feature of this method is that the volume is divided into sub-images at each stage to locally refine T_r and correct for any deformation of the spine. The progressive division into sub-images acts as a method to both refine the local registration estimate for each vertebra to account for the deformation and also to remove the confounding influence of far-off deformed anatomy. Note that the method maintains the advantageous characteristics of the original rigid LevelCheck algorithm, is primarily automatic (i.e., the progression to smaller local regions at each stage does not require additional user input), and is distinct from strictly piece-wise registration (that typically rely on segmentation). The basic parameters in the (single-stage) LevelCheck method were taken from previous studies of the sensitivity of registration performance to parameter values, identifying the nominal values presented above. The key parameters for the msLevelCheck performance are investigated with respect to accuracy and sensitivity as detailed below.

4.2.2.1 Multi-Stage Progression

The key feature of msLevelCheck is that at each subsequent stage, k, the 3D image is divided into multiple 3D sub-images, each focusing on (possibly overlapping) local regions and are independently registered to the p (2D image) using the outputs from the previous stage ($T_{r;k-1}$) to prove a robust initialization for each stage. As illustrated in Fig. 4.3, the first stage of this framework is equivalent to the rigid LevelCheck algorithm, which provides an accurate registration in some portion of the image. At stage 2, independent registrations are performed on sub-images of the 3D CT defined from masked regions about subgroups of vertebral labels (described in Sec. 4.2.2.2). In subsequent stages, the sub-images are further divided to focus on smaller, increasingly local 3D regions until the final stage at which the output registration transforms are used to compute annotation locations on the 2D image. Note that each level of the multi-stage process (i.e., in progressively smaller sub-image regions), a typical coarse-to-fine morphological pyramid was employed as detailed below, but that morphological pyramid should not be confused with the multi-stage process that achieves a globally deformable transformation based on multiple rigid registrations in successively smaller regions of interest. Thus, the multi-stage framework yields a transformation of the annotations from the 3D CT to the 2D radiograph that is globally deformable yet locally rigid to improve the registration accuracy at each annotation.

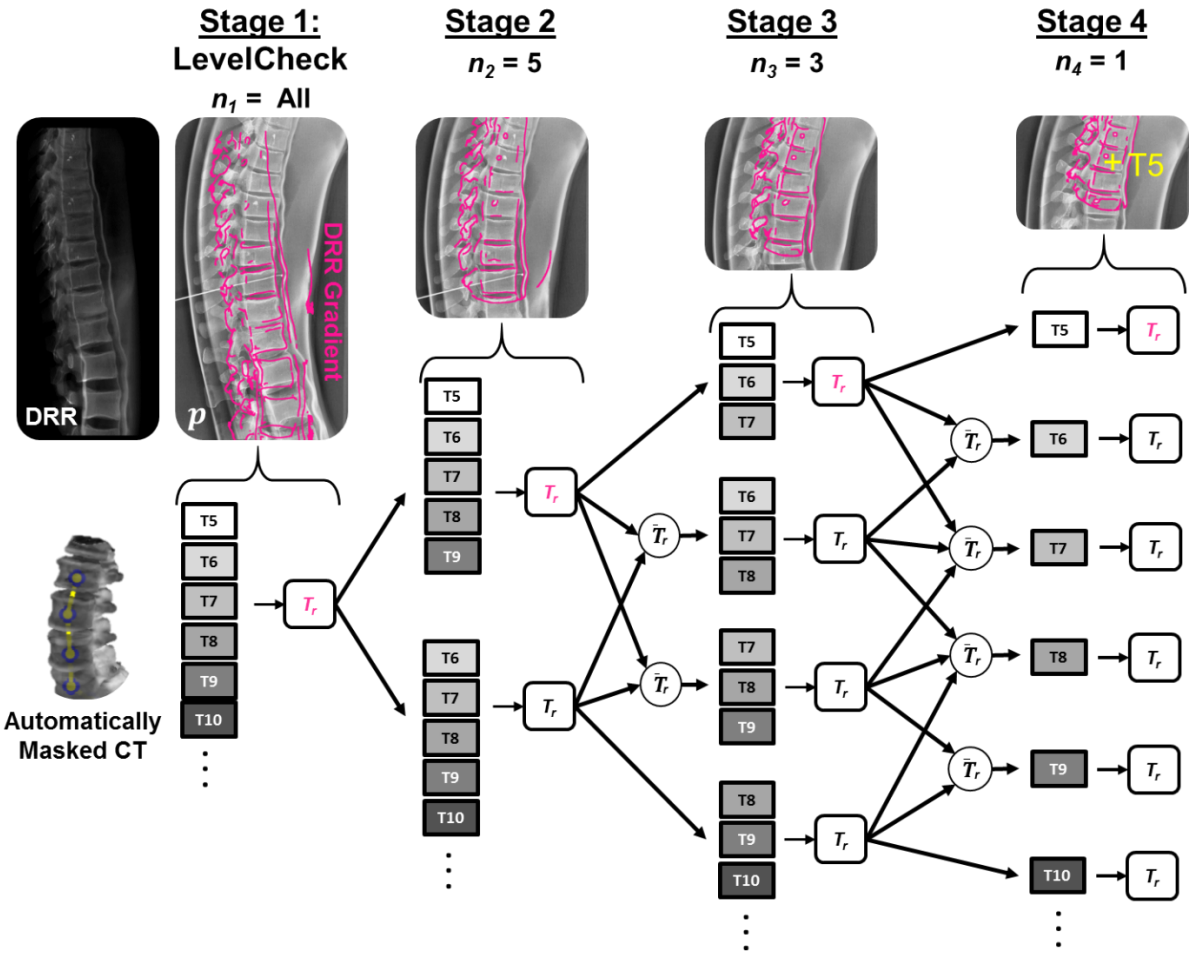


Figure 4.3: Illustration of msLevelCheck using 4 stages with the sub-image size, n_k, for the stages set to {All, 5, 3, 1}. Images along the top show the projection image p with a DRR gradient overlay in magenta, depicting the progression of msLevelCheck along the upper arm of the registration framework for each stage in the multi-stage method. Figure adapted with permission of the publisher from [86].

4.2.2.2 Definition of Sub-Images

To divide the CT into sub-images at each stage, subsets of the 3D preoperative vertebral labels are used to generate 3D binary masks around local regions using the same principle of binary volumetric masking as described in Sec. 4.2.1.1. The size of the sub-images at each stage k is set by n_k, the number of labels chosen to generate each mask (represented as the number of dots in the binary volume mask of Fig. 4.5). Therefore, binary masking provides a

segmentation-free region of interest for various locations along the spinal column [e.g., T5–T6 ($n_k = 2$), T5–T7 ($n_k = 3$), etc.] and, owing to the distance-based mask, may even include spinal levels outside the specified labels. The number of stages (S) and the method for choosing which subsets of the annotations are used to generate each sub-image is customizable to a particular case or application scenario and must be investigated to accommodate the expected degree and type of deformation; however, the method employed below generally follows that of Fig. 4.3. In this method, for each of the S stages, the 3D image is divided into sub-images based on masks that are generated from all adjacent permutations of n_k vertebral labels (i.e., for $n_k = 3$ we have {T5–T7, T6–T8, T7–T9,…} rather than {T5–T7, T8–T10,…}, for example). At the first stage, n_1 is set as the total number of annotated vertebrae ("All") and is identical to the rigid LevelCheck method; at each subsequent stage the value is reduced to perform registration using smaller sub-images.

4.2.2.3 Propagation of Transforms at Each Stage

3D-2D registration is performed independently for each sub-image in the multi-stage framework. Initialization for each sub-image is determined by the $T_{r;k-1}$ outputs of the previous stage from registrations containing the entire region of the current sub-image (as depicted by the arrows in Fig. 4.3). In the scenario where multiple outputs fall into this set, an average over these N_I initialization transformations is used to determine an appropriate initialization. Such an average transformation can be computed by separating the problem into

translation and rotation components. For translation, the mean is computed over the input translation components:

$$\bar{T}_{xyz} = \frac{1}{N_I}\sum_{j=1}^{N_I} T_{xyz}^{(j)} \tag{4.3}$$

Where $T_{xyz}^{(j)}$ is the 3×1 translation vector of the jth T_R. For the average rotation, a quaternion average is computed over rotational components to handle the non-linearity of Euler angles. By representing each of the 3×1 rotation vectors as equivalent 4×1 quaternion rotations, $T_{\eta\theta\phi}^{(j)} \rightarrow q_j$, [106] describes the average of these rotations to be the eigenvector of the matrix M that corresponds to the largest eigenvalue (i.e., U_1, the first column of U when S is a decreasing diagonal matrix and USU^t is the eigen-decomposition of M):

$$M = \sum_{j=1}^{N_I} q_j {q_j}^t = USU^t \tag{4.4}$$

Following this decomposition, the average quaternion rotation is transformed back into Euler angles, ($U_1 \rightarrow \bar{T}_{\eta\theta\phi}$), and $\bar{T}_r = \left[\bar{T}_{xyz}^t, \bar{T}_{\eta\theta\phi}^t\right]^t$ is used to initialize the subsequent registration.

4.2.2.4 Scaling optimization parameters

The accuracy is expected to gradually improve as the multi-stage registration progresses, and registration parameters are accordingly adjusted to a finer range and scale. As the transformation estimate approaches the solution at each stage, parameters governing the search range (*SR*, as outlined in Sec. 4.2.1.4) are scaled to better suit the smaller region of interest and improve registration runtime. In terms of decreasing *SR*, the parameters of T_r governing the translation direction z_r (corresponding to magnification) and the three rotations were reduced

to relatively small empirically-determined fixed values at each stage to cater to the maximum amount of expected deformation. On the other hand, the remaining two translation parameters x_r and y_r (most directly corresponding to $[u,\ v]$ on the detector) demonstrated greater variability across stages, and thus were reduced in an adaptive manner according to the variation of output poses from the previous stage. The search range, $SR_{x,y}$, for these parameters consisted of the addition of two components: (i) a fraction, f_k , of the intervertebral distance (IVD, i.e., the computed mean distance between adjacent vertebral labels on the detector computed from the estimated projected labels of the previous stage); and (ii) an adaptive term, D_a, that extends the SR by the standard deviation among label positions computed from the multiple T_r poses used for the initialization — thus SR is smaller if the outputs of the previous stage are in agreement. We selected IVD as a reference based on the finding that registration following stage 1 tends to be accurate within the range of one vertebra; therefore, choosing a search range based on IVD provides a consistent method to constrict the search range (via reducing f_k) in a manner that normalizes effects of patient size and vertebra type (i.e., cervical/thoracic/lumbar).

$$SR_{x,y}(f_k) = \frac{z_r}{SDD}(f_k \times IVD + D_a)$$

$$\text{where} \quad D_a = \sqrt{\frac{1}{n_k N_I}\sum_{i=1}^{n_k}\sum_{j=1}^{N_I}\left\|d_{ij} - \bar{d}_i\right\|^2} \tag{4.5}$$

$$\text{and} \quad IVD = \frac{1}{(n_k - 1)N_I}\sum_{j=1}^{N_I}\sum_{i=1}^{n_k-1}\left\|d_{ij} - d_{i+1,j}\right\|$$

The term D_a is the standard deviation of the projected label positions on the detector, $d_{ij}(u, v)$. To compute D_a, the N_I initialization poses $[\ T_r^{(j)}]$ are used to project each of the n_k labels

included in the mask for current registration to achieve d_{ij} (the projection of label i onto the detector using $T_r^{(j)}$). The standard deviation is then computed by calculating the distance of each d_{ij} to the centroid location for its associated vertebra, $\bar{d}_i$ (mean across j of d_{ij}). This term is added to the fraction of the IVD (i.e., $f_k \times IVD$) and scaled by the inverse of the current magnification estimate (z_r/SDD) to approximate this distance in the CT world coordinates. The search range $SR_{x,y}(f_k)$ therefore provides an increasingly smaller search range (by reducing f_k at each stage) that is extended adaptively based on the agreement among the poses in the previous stage. With this smaller SR and an improved initialization estimate, optimization parameters *MS* and λ can be relaxed without diminishing performance. Therefore, following stage 1, to improve computation time and reduce GPU memory, *MS* and λ were reduced to 25 and 100, respectively, before noticeable stochastic effects were observed in the CMA-ES optimizer.

4.2.2.5 Enhancing structural image features

Each stage in the method facilitates finer registration accuracy and exploits increasingly fine detail of anatomical structures in the underlying images. To achieve a finer level of detail, the downsampling of p is reduced (by decreasing a_{pix}) along with the kernel width σ (characteristic width of the Gaussian smoothing kernel) for the image gradient calculation when computing the metric GO. A parameter sensitivity study that tested 100 variations of a_{pix} and σ for stage 1 registration indicated stable performance near 2 mm for both parameters. Following stage 1, the choices for a_{pix} and σ were incrementally reduced to the final stage value of 1.5 mm and 1.25 mm, respectively, based on empirical tests in a small number of samples

and recognizing limitations in GPU memory (noting that a_{pix} reduction yields a quadratic factor increase in GPU memory use). As a further step to improve memory efficiency, the p image is cropped to contain only the region that is defined by the search range and sub-image extent of the current registration. Following the first stage, adaptive histogram equalization is applied to the radiograph to locally enhance the contrast and thereby accentuate structures that may otherwise fall beneath the gradient threshold applied during GO calculation, an effect that becomes increasingly likely as the impact of noise rises due to the reduction in down-sampling and gradient kernel width.

4.2.3 Experimental Methods

4.2.3.1 Single-stage registration with sub-image extent n_1

A sensitivity study was performed to investigate robustness under the scenario of using only one stage in which the 3D CT image is immediately divided into sub-images of size n_1 (ranging from 1 to 7 vertebrae). For example, in scenarios for which the structure of interest is just a single vertebral target level, it may be of interest to directly register a small sub-image about that level. We therefore investigated the question of how many vertebrae are necessary for a successful registration, particularly in cases of poor initialization.

The robustness of single-stage registration for such small sub-images was evaluated in an IRB-approved retrospective clinical data set of 24 patients undergoing thoracolumbar spine surgery, consisting of 24 CT images and 61 intraoperative radiographs. Preoperative CT included data from three scanner manufacturers (Siemens Healthcare, Erlangen, Germany;

Toshiba Corporation, Tokyo, Japan; and GE Healthcare, Little Chalfont, UK) with scan techniques ranging from 120−140 kVp, 80−660 mAs, and 0.24−3.00 mm slice thickness. Intraoperative radiographs were all acquired with a mobile radiography system (DRX-1, Carestream Health, Rochester, NY, USA) with pixel dimensions 0.14 x 0.14 mm^2. Binary Volumetric masks were automatically generated with the number of adjacent vertebrae (n_1) ranging from 7 down to 5, 3, and 1, centered on a central vertebra in the radiograph. Registration was performed using the rigid LevelCheck algorithm (full search range) with the stage 1 parameters in Table 4.2, as described in Sec. 4.2.1.

Registration accuracy was evaluated in terms of the projection distance error (PDE) for each label at the detector, defined for the i^{th} label as:

$$PDE_i = \|t_i(u,v) - d_i(u,v)\|_2 \tag{4.6}$$

which is the distance between the projected CT label $[d(u,v)]$ to the ground truth label $[t(u,v)]$. The ground truth position (approximately located at the vertebral body centroid) was manually defined in the radiograph by an expert neuroradiologist. Successful localization involves registration of a label within the bounds of the vertebral body — approximately 15 mm radius for a thoracolumbar level — corresponding to 22.5 mm at the detector for magnification M = 1.5. Therefore, registration failure was conservatively defined as cases for which the mean PDE among projected labels was greater than 20 mm. Intra-user variability in centroid identification was analyzed (and as shown below, found to be the main source of variability in the resulting PDE). Taking again M = 1.5, and the intra-user variability in centroid definition σ_{CT} = 2 mm in the CT image, and σ_{rad} = 2 mm in the radiograph, the resulting error is $\sigma_{def} = \sqrt{(1.5^2)(2^2) + (2^2)}$ = 3.6 mm. This level of intra-user variability

is sufficient for purposes of level definition (i.e., < ~20 mm failure criterion). Further, we investigated the influence of anatomical deformation (changes in spinal curvature) when the region being registered is large (for example, $n_1 \geq 7$), noting that even in cases of successful global registration (mean PDE < 20 mm), there may be individual labels exhibiting large PDE, as illustrated in Fig. 4.1C; therefore, the maximum PDE was also examined.

4.2.3.2 Multi-stage framework determination

The msLevelCheck method does not impose explicit constraints on the projected labels, only a $\overline{T_r}$ calculation that, coupled with a reduced search range at each stage, acts as an implicit regularization on adjacent vertebral motion. To investigate the degree of regularization that is necessary (in terms of both the number of stages and the sub-image size at each stage), seven potential frameworks for msLevelCheck were tested on a challenging case from the clinical data set. Thus, the goal of this study was to determine a sufficient framework that minimized the number of stages necessary while still being able to accurately correct for the magnitude of deformation that can be expected in spinal procedures. The frameworks presented in Table 4.1 represent a variety in both the depth and structure of possible multi-stage registration trees: from short trees (e.g., {All, 1}) for which the sub-images are immediately divided into small segments to deep trees (e.g., {All, 5, 4, 3, 2, 1}) for which the sub-image division is incremental.

Table 4.1**:** Registration frameworks considered for msLevelCheck. Framework notation for n_k over a number of stages (S) is denoted in { } brackets, with 'All' denoting all vertebrae within the radiographic field of view. For example, {All, 5, 3, 1} denotes a four-stage framework in which the registration is computed for all vertebrae (as in the basic LevelCheck algorithm), followed by 5, 3, and finally each (1) single vertebrae. Performance of each framework is shown in Fig. 4.6.

	Framework #						
	1	**2**	**3**	**4**	**5**	**6**	**7**
$\boldsymbol{S}$	2	2	2	3	3	4	6
$\boldsymbol{n_k}$	{All, 1}	{All, 2}	{All, 3}	{All, 3, 1}	{All, 4, 1}	{All, 5, 3, 1}	{All, 5, 4, 3, 2, 1}

Following these choices for S and n_k, the remaining parameters to be selected include, f_k, a_{pix} and σ. By analyzing the clinical dataset, it was seen that (within the region spanned by the radiograph) deformation of the spine tended not to exceed a distance proportional to approximately half of a vertebral body length. Since IVD includes the distance of a full vertebral body plus the intervertebral disk space, to account for the expected level of max deformation, f_2 was set to be 0.4 for each tested framework. For the following stages, it is expected that the initialization is increasingly nearer to the solution; thus, f_k was reduced incrementally, roughly in proportion to the decrease in the sub-image size. As described in 2.5.3, a_{pix} and σ were similarly decreased based on n_k from 2 mm (at stage 1) to 1.5 mm (at n_k = 1) for a_{pix} and 2 mm to 1.25 mm for σ. A table of parameter values for framework 6 is provided in Table 4.2.

The registration of CT labels to the radiograph was repeated 5 times for each framework, and the PDE was analyzed to determine a suitable framework for subsequent studies.

Table 4.2: Summary of nominal parameters in the msLevelCheck algorithm, framework 6.

	n_k	a_{pix} (mm)	σ (mm)	***SR:*** ± [***x, y, z, η, θ, ϕ***] (mm, mm, mm, °, °, °)	***MS***	λ
Stage 1 (rigid)	All	2.00	2.00	[100, 200, 75, 15, 10, 10]	50	125
Stage 2	5	1.75	1.50	[$SR_{x,y}(0.4)$, $SR_{x,y}(0.4)$, 20, 10, 5, 5]	25	100
Stage 3	3	1.75	1.50	[$SR_{x,y}(0.2)$, $SR_{x,y}(0.2)$, 10, 5, 5, 5]	25	100
Stage 4	1	1.50	1.25	[$SR_{x,y}(0.15)$, $SR_{x,y}(0.15)$, 10, 5, 5, 5]	25	100

4.2.3.3 Multi-stage registration in phantom

Evaluation of msLevelCheck in the presence of deformation was performed using a Sawbones® spine phantom (Pacific Research Laboratories, Inc., Vashon Island, WA, USA) within a flexible bulk holder simulating adjacent soft tissue as shown in Fig. 4.4. A CT scan emulating the preoperative pose was acquired with the phantom lying flat (as in Fig. 4.4A–B, scanned on Toshiba Aquilion One, 120 kVp, 400 mA, Bone standard reconstruction, 0.625 x 0.625 x 0.5 mm^3 voxel size). Spinal deformation analogous to that of an arched Wilson operating table was simulated using foam core inserted below (anterior to) the spine as in Fig. 4.4C. Six inserts varying in thickness from 4.9 to 10.1 cm gave a total of 7 curvatures, ranging from flat to strongly kyphotic. For each of the 7 levels of deformation, lateral projection images were acquired using a mobile radiography system (DRX-1, Carestream Health, Rochester, NY, USA) analogous to the images acquired in intraoperative spine level localization, spanning a region of the thoracolumbar spine from roughly T5 to L2. The true location of each vertebral level was manually defined in the CT and each projection image by a single observer (an engineer familiar with the relatively simple anatomy in this phantom).

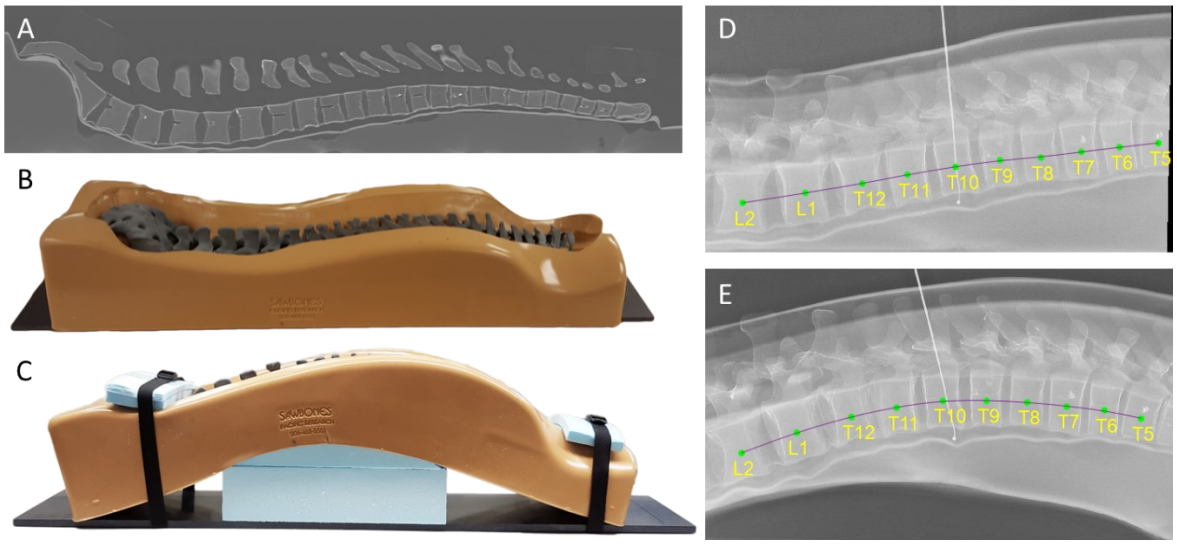


Figure 4.4: Investigation of spinal deformation in phantom. (A) Sagittal CT slice of the (B) spine phantom lying flat. (C) Photograph of the spine phantom with maximally induced curvature. (D) Lateral radiograph with vertebral levels overlaid of the phantom lying flat, as in (B). (E) Lateral radiograph of the phantom with maximal deformation, as in (C), overlaid with level labels. Figure adapted with permission of the publisher from [86].

For this study, the four-stage msLevelCheck framework illustrated in Fig. 4.3 was used, with the number of vertebrae in each mask at each stage set to n_k = {All, 5, 3, 1} (i.e., framework 6 in the previous study). This hierarchy was chosen for this setting following the experiment of Sec. 4.2.3.2 where it proved robust in balancing the tradeoff between solving large deformations and avoiding local optima. The parameters for this framework are detailed in Table 4.2, with parameter choice as described in Sections 4.2.2 and 4.2.3.2.

The msLevelCheck registration of "flat" CT to "deformed" radiograph was repeated 5 times for each of the 7 deformation cases to test the robustness due to the stochasticity of the CMA-ES optimizer. Analysis of registration accuracy was performed by examining the distribution of PDE values for the 7 deformation cases and comparing the performance to the rigid (conventional single-stage LevelCheck) registration. Statistical significance in the difference between rigid and msLevelCheck results was analyzed using a Wilcoxon signed-rank test.

4.2.3.4 Multi-stage registration in clinical data

The msLevelCheck method was further tested using a subset of the clinical data described in Sec. 4.2.3.1. The most severe deformation cases among the 61 radiographs were selected by analyzing the rigid registration result (for which the vertebrae in the center of the radiograph was typically well registered) and computing the increasing trend in PDE for vertebrae superior and inferior to the central vertebra (i.e., cases exhibiting greater PDE at superior and inferior extrema). From these data, 7 radiographs from 5 patients were selected as exhibiting the most severe deformation. Using the same four-stage method detailed in Fig. 4.3 and Table 4.2, the rigid and msLevelCheck methods were evaluated by examining the mean and maximum PDE.

4.2.3.5 Comparison to piecewise rigid registration

To test the performance of msLevelCheck to alternative methods, we performed a comparative study to a piecewise rigid approach. Note that the msLevelCheck method is segmentation-free, whereas the piecewise rigid method assumes a reliable segmentation. The piecewise rigid registration was performed by first segmenting individual vertebral bodies in the CT of the spine phantom. Segmentation was accomplished using the active contour method implemented in ITK Snap [107]. A central segmented vertebra was initialized near solution (error to within half of a vertebral body length) for the maximum deformation radiograph case, and 3D-2D registration was performed for the individual segmented vertebrae. Adjacent vertebrae were recursively registered using the output transformation from a previously registered adjacent vertebra as an initialization. Search range constraints were imposed to prevent one-level vertebral “jumps” in registration. Registrations using msLevelCheck and the

piecewise rigid method were repeated 5 times each on this maximum deformation spine phantom case. Performance was evaluated in terms of PDE (mean and standard deviation), hypothesizing comparable performance between msLevelCheck and piecewise rigid, but recognizing the advantage of the former operating without segmentation.

4.3 Results

4.3.1 Single-stage registration with sub-image extent n_1

Figure 4.5 shows the performance of single-stage rigid registration as a function of sub-image size, n_1. This single-stage algorithm is equivalent to the previously described LevelCheck algorithm for various choices of n_1 and thus examines performance as the region of support decreases. Fig. 4.5A shows examples of the volumetric mask size, DRR, and labeled projection for variable n_1. As shown in Fig. 4.5B, the failure rate increases sharply for fewer vertebrae within the mask, indicating that reliable registration benefits from including longer extent (more vertebrae) in the binary volume mask definition, providing a larger field of view for the registration and improved robustness against local minima. As shown on the second vertical axis of Fig. 4.5B, however, spinal deformation causes some regions within the larger field of view to align poorly (typically at the superior / inferior ends of the image) and exhibits an increase in maximum PDE. Therefore, the LevelCheck algorithm benefits from a larger field of view (increased n_1) but can suffer from anatomical deformation, motivating the multi-stage registration framework analyzed below.

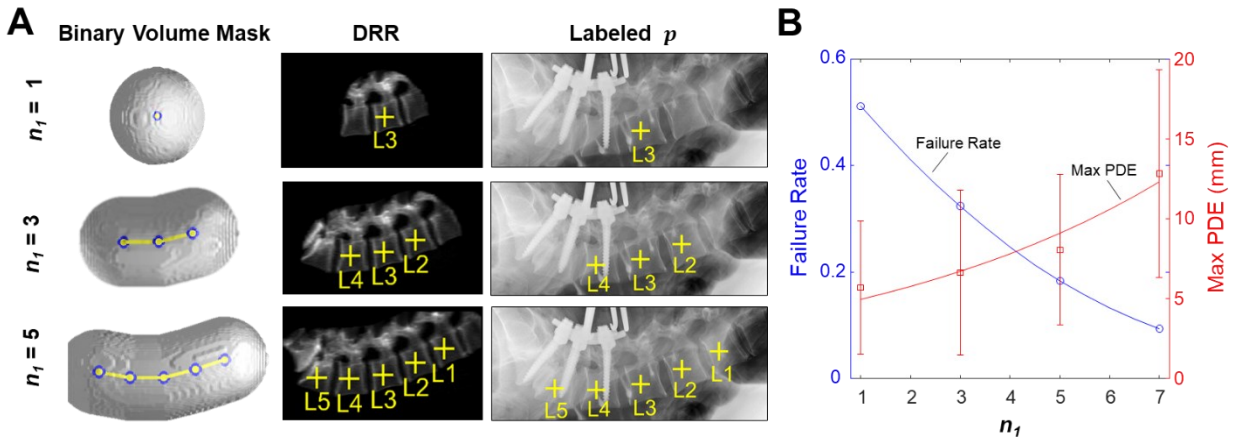


Figure 4.5: Sensitivity to the number of vertebrae included in single-stage registration evaluated in 61 clinical radiographs. (A) Examples show n_1 = 1, 3, and 5 vertebrae, each with a 50 mm binary volume mask. (B) Failure rate and maximum PDE measured as a function of n_1. The observation that smaller mask size reduced the max PDE motivated development of the msLevelCheck method to provide both robust global registration (via the initial stages) and more accurate registration local to each vertebra (via the end stages). Figure adapted with permission of the publisher from [86].

4.3.2 Multi-stage framework determination

Figure 4.6 compares the performance of the various multi-stage frameworks shown in Table 4.1. Two types of error mode are evident. For the shorter (S=2) trees, a broader, more uniform distribution of error is seen, with errors widely distributed over a range of ~15 mm PDE. In the deeper trees, we observed PDE concentrated near ~2.5 mm; however, a fairly large number of failures (PDE >20 mm) were evident for the 3-stage frameworks. The 4-stage framework {All, 5, 3, 1} provided low PDE with the fewest outliers, and the deeper 6-stage framework did not provide further significant improvement. Therefore, the {All, 5, 3, 1} tree was selected as the nominal framework for msLevelCheck in subsequent studies.

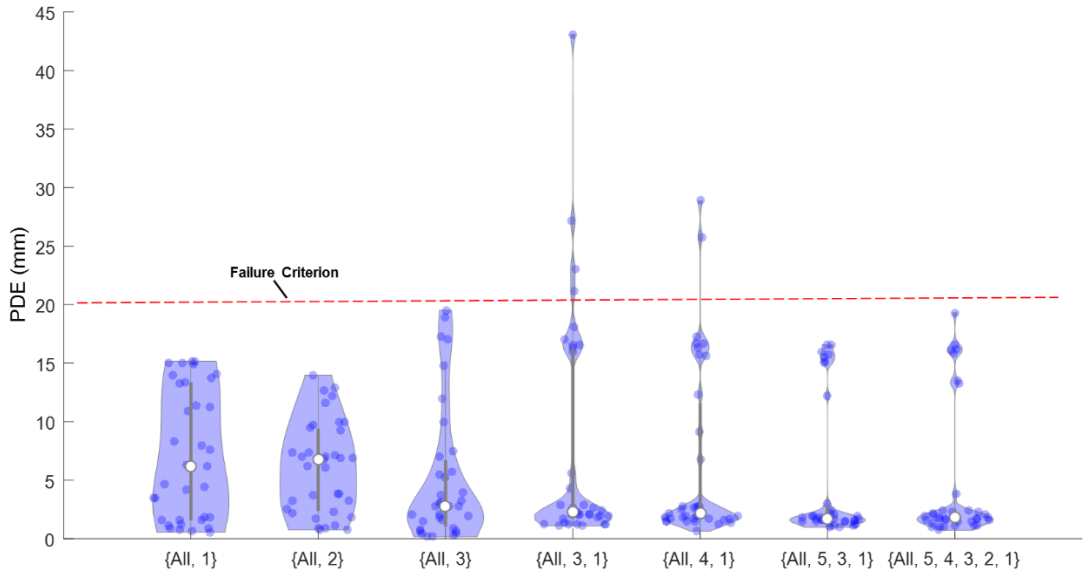


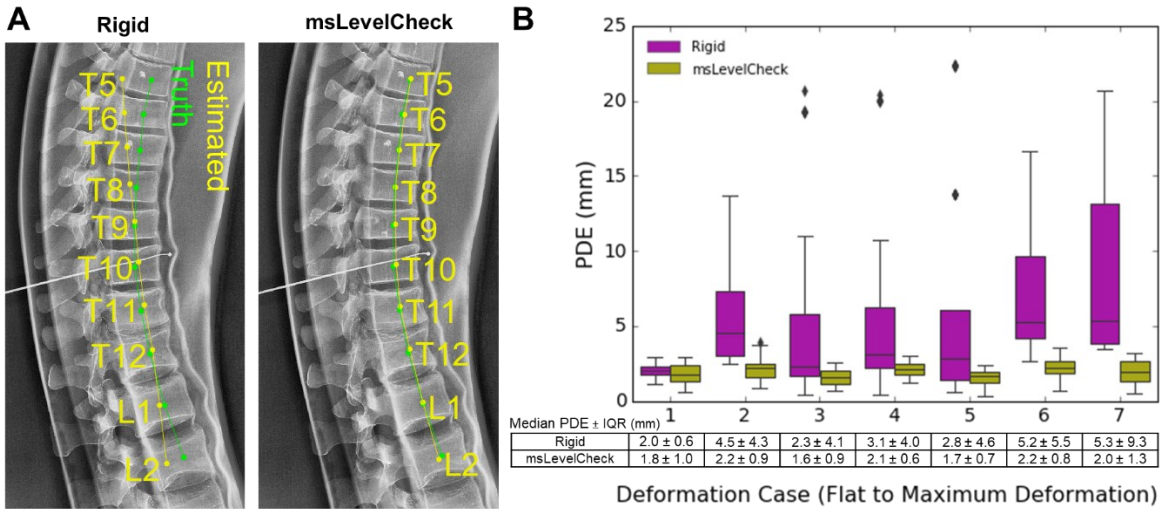
Figure 4.6: Comparison of various multi-stage frameworks listed in Table 4.1. Violin plots indicate the distribution of PDE for the registered labels in each framework. Figure adapted with permission of the publisher from [86].

4.3.3 Multi-stage registration in phantom

Figure 4.7 summarizes the msLevelCheck performance measured for different degrees of spinal deformation. Figure 4.7A shows an example for the strongest deformation (case 7), where the rigid approach (stage 1) tended to align well in the central region of the radiograph but decreased in accuracy at the inferior and superior ends of the image. The msLevelCheck method, however, was able to accurately map all vertebral labels by incrementally focusing alignment on sub-image regions, improving alignment locally at each stage.

Figure 4.7B quantifies the performance improvement in terms of PDE for the seven cases of increasingly strong deformation. For the rigid method, although the median PDE is ≤ 6 mm for all cases, the interquartile range (IQR) and frequency of outliers increased steadily with stronger deformation. Despite such challenge, the msLevelCheck method was unaffected,

maintaining median and IQR in PDE across the full range of deformation examined in this study. The distribution in PDE for msLevelCheck showed a statistically significant improvement ($p < 0.001$) for each case, except for case 1 (no deformation, where both methods performed well). With respect to outliers and maximum PDE, the rigid method showed maximum PDE = 22.4 mm (in case 5), whereas msLevelCheck gave a maximum PDE of 3.9 mm (in case 2), which is below the value indicative of failure (~20 mm).



	1	2	3	4	5	6	7
Rigid	2.0 ± 0.6	4.5 ± 4.3	2.3 ± 4.1	3.1 ± 4.0	2.8 ± 4.6	5.2 ± 5.5	5.3 ± 9.3
msLevelCheck	1.8 ± 1.0	2.2 ± 0.9	1.6 ± 0.9	2.1 ± 0.6	1.7 ± 0.7	2.2 ± 0.8	2.0 ± 1.3

Figure 4.7: Registration accuracy for the msLevelCheck method under various degrees of deformation (spinal curvature). (A) Illustration of registration for the single-level rigid and msLevelCheck methods for the case of strongest deformation (case 7). (B) Boxplots depicting the distribution of PDE for both registration methods for the 7 deformation cases (cases 1–7, indicating increasing degree of deformation) along with the tabulated numerical values for median PDE and IQR. Figure adapted with permission of the publisher from [86].

4.3.4 Multi-stage registration in clinical data

Figure 4.8 compares the performance of the rigid and msLevelCheck methods applied to clinical data. Figure 4.8A shows an example case in which the spine was more lordotic in the intraoperative radiograph than the preoperative CT, consistent with typical patient positioning on a Jackson table. The rigid method is seen again to provide good registration at the center of the image, but to lose accuracy at superior and inferior extrema. Figure 4.8B depicts the

distribution in mean PDE (and Fig. 4.8C the maximum PDE) aggregated over all cases in the clinical data set, demonstrating a statistically significant improvement ($p < 0.001$) in both mean and maximum PDE for msLevelCheck. Overall, the average PDE improved from 8.1 mm with the single-stage rigid method to 4.6 mm with msLevelCheck. More importantly, the maximum PDE was reduced from 32.0 mm for the single-stage rigid method to 18.6 mm for msLevelCheck. It bears repeating that cases selected in the clinical study were those exhibiting the most severe deformation, and while the single-level rigid approach resulted in 6.5% of labels falling outside the 20 mm failure criterion (i.e., outside or near the boundary of a given vertebra) in a manner that may diminish utility of the algorithm, the msLevelCheck method registered all vertebrae within the acceptable range. Note also that the algorithm maintained other desirable aspects of the original LevelCheck method, such as robustness against the presence of interventional tools (as illustrated in Fig. 4.8C).

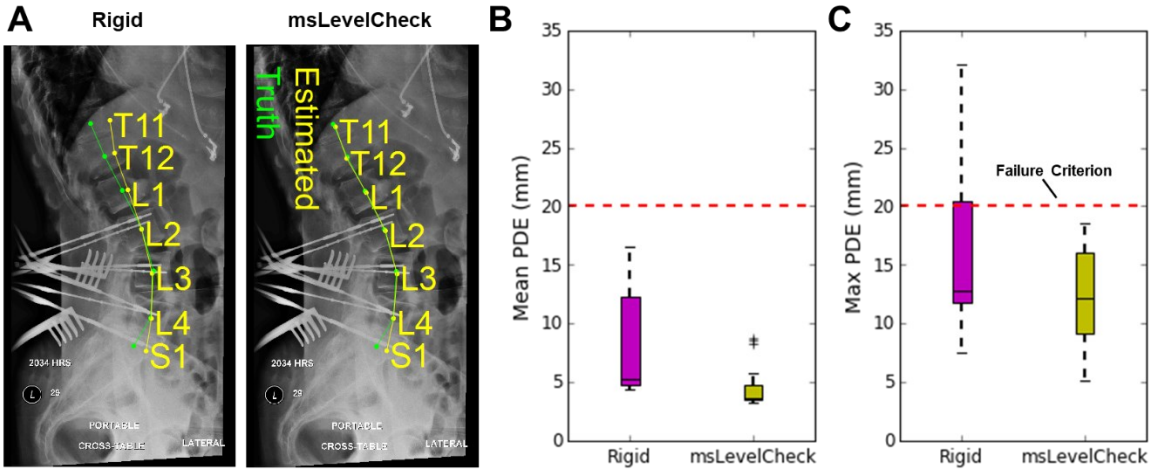


Figure 4.8: Registration accuracy for msLevelCheck in clinical data. (A) Example case showing single-level rigid registration and msLevelCheck output for a case exhibiting an increase in spinal lordosis in the radiograph compared to preoperative CT. Distribution the mean (B) and maximum (C) PDE pooled over cases in the clinical dataset, showing msLevelCheck to improve registration accuracy and recover from cases that might be considered a registration failure. Figure adapted with permission of the publisher from [86].

4.3.5 Comparison to piecewise rigid registration

Figure 4.9B examines the distribution in PDE for the piecewise rigid method in comparison to msLevelCheck (for the {All, 5, 3, 1} framework), yielding PDE = (2.5 ± 1.9) mm and (2.0 ± 1.4) mm, respectively (median PDE ± IQR). Both methods performed successfully (PDE < 20 mm), and there was no statistically significant difference between the two distributions (paired *t*-test p-value > 0.05). This indicates that msLevelCheck performed at least as well as the piecewise rigid solution but with the benefit of not requiring explicit segmentation involved in the piecewise rigid method.

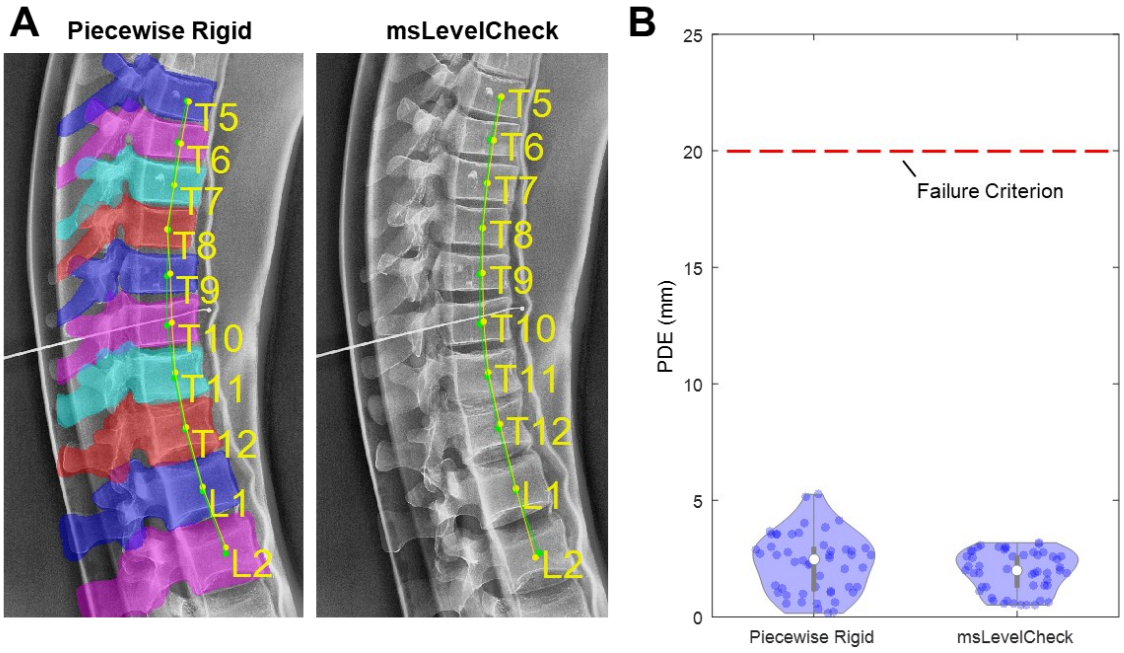


Figure 4.9: Comparison of performance for piecewise rigid and msLevelCheck. Results are shown for the case of maximum deformation in the spine phantom. (A) Illustration of registration for the piecewise rigid (overlaid with the projections of the requisite vertebrae segmentations) and msLevelCheck methods. (B) Violin plots show the distribution of PDE for the registered labels in each method, with median PDE shown as a solid white circle, upper and lower bounds given by the max and min PDE, and 50 individual sample points shown therein. Figure adapted with permission of the publisher from [86].

4.4 Conclusion

In this chapter, we presented a multi-stage 3D-2D registration algorithm (msLevelCheck) for mapping label annotations (e.g., vertebral labels or other point features demarked in preoperative 3D images as part of existing clinical workflow) to intraoperative radiographs under conditions of strong anatomical deformation. The multi-stage approach amounts to 3D-2D registration that is locally rigid (at progressively finer scales) and yet globally deformable (with respect to mapping of label annotations from the 3D image to the radiograph). The latter point bears repeating: the method does *not* constitute a deformable registration of the *image*; rather, it produces a series of rigid transformations by which point annotation *landmarks* are transformed independently, and thus *deformably* with respect to the underlying image. The progressive multi-stage registration framework is shown to be necessary in Sec. 4.3.1 where we see a high rate of registration failure when the sub-images are immediately broken up into small sub-images. Further, Sec. 4.3.2 indicates that a 4 stage {All, 5, 3 ,1} framework provides a sufficient rate of sub-image reduction for this application.

A recent study evaluated the clinical utility of LevelCheck [85] and showed that a clinician's confidence in target localization could be diminished when some labels (even if far from the target level) were placed near or outside the periphery of a vertebral body. The msLevelCheck method addresses this concern, showing accurate registration of all vertebral levels (2.9 mm median PDE, 3.8 mm IQR, and 0% failures) over a broad range of deformation in clinical data.

The registration runtime is necessarily increased for the multi-stage framework. Previous work showed the original LevelCheck method to run in ~20 s [84], and runtime as long ~60 s was said to be acceptable within clinical workflow for purposes of decision support in level

counting [85]. In the current work, the framework was implemented in a simple serial form in which the runtime increases in proportion to S (the number of stages) and V (the number of vertebral labels), implying an increase in runtime by a factor of $\sim 4V$ for the 4-stage {All, 5, 3, 1} framework of Fig. 4.3. However, because each stage contains multiple independent registrations, these registrations within each stage can be parallelized; therefore, in a parallel implementation the increase in runtime is instead proportional to S. Moreover, since SR, MS, and λ are reduced following stage 1, the runtime associated with each registration in these later stages is faster than the first stage (original LevelCheck method). In terms of the number of function evaluations that must be completed along each parallel string (i.e., each sub-image registration), ~250,000 function evaluations were typically required in the first stage, whereas ~110,000 function evaluations were typical for the following stages. A more optimal implementation of the 4 stage method to be developed in future work would therefore scale the runtime by a factor of ~2.3 compared to the original LevelCheck algorithm, amounting to ~50 s.

Apart from anatomical deformation, factors that are known to challenge the original LevelCheck method (delineated in Lo *et al.* [108]) may similarly challenge msLevelCheck. These include poor radiographic image quality (e.g., large patients and/or poor radiographic technique), poor CT image quality (e.g., thick slices), high density of surgical instrumentation, and gross anatomical mismatch (e.g., corpectomy). To better address these challenges, the msLevelCheck method could be extended to take better advantage in regularizing the entire series of local registration outputs accrued across each stage. By analyzing the trend in output pose across stages and among neighboring registrations, outliers in individual registrations

could be detected and trapped — either by retreating to the output of a previous stage or interpolating across the pose estimates for adjacent regions.

Throughout this chapter we observed multiple instances in which the statistical model from Chapter 3 provided important guidance for the development of the registration method. In this application, not only did content mismatch play a role due to the presence of soft tissue and surgical instrumentation in the radiograph, but we also observed global spine deformation acting as a source of noise. By understanding these sources of noise in the context of Chapter 3, we demonstrated that gradient-based similarity metrics and sub-image cropping to effectively compensate for these confounding influences and achieve accurate vertebral labeling. It is also important to note factors such as proper initialization and the impact of local optima, which are not directly addressed by the statistical model, but are important considerations when developing a registration method. For instance, the motivation for the number of stages came primarily from the need to provide a close initialization to each subsequent stage — particularly as we noted in Fig. 4.5C that registration failure becomes increasingly likely as the sub-images get smaller when using only rough initialization.

In the next chapter, we depart from the context of rigid registration to understand the factors that influence registration performance in CNN-based deformable registration, particularly with respect to statistical differences between the training and test dataset.

Chapter 5: Learning-Based Deformable Image Registration: Effect of Statistical Mismatch Between Train and Test Images

5.1 Introduction

In the previous chapters, we presented both theory and implementations for rigid image registration, particularly in the context for understanding how anatomical deformation and factors affecting image quality contribute to registration performance. In this chapter, we will examine CNN-based deformable registration, which presents two interesting points that should be considered in the context of the previous chapters: (1) resolving soft-tissue deformation is now the goal of the registration task and must be considered as true-signal content (rather than a noise source) when extending the statistical model to deformable registration; and (2), not only should we consider the image quality of the images that are being registered, but also that of the data used to train the CNN model. Below, we investigate these questions while keeping

in mind the importance for CNN-based methods to generalize to data not observed during training.

CNNs are increasingly being investigated as a method for deformable registration in medical imaging [109]–[114] due to their fast runtime and ability to learn complex functions without explicit physical models. Compared to conventional methods for image registration such as B-spline free-form deformation [115] and variations on diffeomorphic registration [28]–[30], [116], CNN-based methods are not only generally much faster [109], [112] but also provide a parameter-free, non-iterative interface for achieving registration. However, a recurring question associated with CNN methods is the generalizability of the model beyond the data presented in the training set. This question is commonly addressed by dividing the data into train and test sets and performing cross-validation studies. However, the random sampling associated with this method enforces that the train and test data have the same population statistics, which could be unrealistic for various application scenarios in medical imaging.

A clear example of this effect was shown by Eppenhof *et al.*[112], who performed CNN-based registration on pulmonary CT images and examined two separate data sets, DIR-Lab [117], [118] (images acquired using a GE Discovery ST PET/CT scanner) and CREATIS [119], [120] (images acquired using a Philips 16-slice Brilliance Big Bore Oncology Configuration). The authors reported that when the network was trained on DIR-Lab images alone, the results on cross-validated studies were optimistic compared to the results obtained by testing on the CREATIS dataset. One explanation for this deficit is that the network — having only been trained on one dataset with particular image statistics (e.g., noise and resolution characteristics) — was not able to fully generalize to the statistical mismatch that

existed between the datasets from two different scanner manufacturers, each with distinct acquisition and reconstruction protocols and, therefore, spatial resolution and noise characteristics.

Statistical mismatch — by which we mean a difference in some statistical characteristic of the image data, including first-order statistics (e.g., signal power and spatial resolution) and second-order statistics (e.g., NPS) — is of particular concern in medical imaging, where small data sets are unlikely to capture large variations observed in the population. For example, even within a single anatomical region and the relatively reproducible modality of CT imaging, first- and second-order statistics can vary widely based on the scanner manufacturer, scanning protocol (e.g., dose or beam energy), reconstruction protocol, and post-processing technique. Training with all possible variations encountered in practice would be impractical and require unrealistically large training sets. Therefore, in scenarios with known statistical mismatch from the training set, the user opts either to retrain the network or assume the network would reasonably generalize to the test data. For example, when the statistical characteristics of the data are substantially mismatched (e.g., application to MR images using a model trained on CT images), the need to retrain or apply transfer learning is clear. However, with known differences in first- and second-order image statistics between the training and test data, (e.g., training on high-dose and testing on low-dose data), generalizability of the model may be possible with a clear understanding of the extent and the limitations of generalizability.

In this chapter, we will use a classical CNN model for deformable image registration to examine the effect of statistical mismatch in image noise, spatial resolution, and deformation magnitude. By training the network under a variety of statistical conditions in simulated image data, we can observe the performance (registration accuracy) as the statistical characteristics

of the test data deviate from those of the training data. In doing so, we will further extend the statistical model presented in Chapter 2 to compare the experimental performance with the CRLB for deformable image registration. We will validate the findings of these experiments by deploying the networks (trained on simulated image content alone) on anatomical image content. Finally, we will consider how the statistical model and the findings from this chapter may be applied in development of multi-modal deformable registration method that relies on image synthesis.

The work appearing in this chapter was reported in the following conference proceeding and journal papers: (M.D. Ketcha et al., *Proc. SPIE Medical Imaging,* 10949, 2019) [121] and (M.D. Ketcha et al., *J. Med. Imaging.,* 6(4), 2019) [122].

5.2 Background and Theory

5.2.1 CNN-Based Deformable Registration Techniques

CNN-based methods for performing deformable registration are generally grouped into three categories with respect to training: supervised, semi-supervised, and unsupervised. Supervised methods rely upon accessibility to ground-truth, dense displacement vector fields, where the error between the predicted and known displacement fields is directly minimized. The ground truth displacement fields are typically generated by interpolating a displacement field based on corresponding landmarks [123] or applying known displacement fields to simulate deformation [112], [124]. Semi- or un-supervised techniques, on the other hand, still predict an output displacement vector field; however, the network further incorporates a spatial transformer such that, during training, an image similarity measure may be optimized, [110],

[113], [114] which can be paired with deformation-field regularization to yield cost functions similar to that of the conventional registration techniques [125]. Furthermore, another active area of research considers generative adversarial network (GAN) [126] methods (which can be performed in both supervised and unsupervised settings), where the training comprises alternating optimization of a generator network (i.e., a deformation field estimator) and a discriminator network (e.g., which predicts whether or not the registration comes from ground truth or the generator) [109].

5.2.2 Statistical Evaluation of Deformable Image Registration

In Chapter 2, we presented a model relating image quality characteristics (namely, MTF and NPS) to the CRLB for *rigid* image registration (Eq. 2.12). By approximating deformable registration as independent, locally-rigid translation-only registrations at each pixel, we may perform a sliding window computation of the rigid CRLB over the image to determine an approximate deformable CRLB at each pixel location, yielding a spatially varying "CRLB Map" (Fig. 5.1B). From this map, we observe that regions with high gradient content have a reduced CRLB and are thus more reliable regions for driving registration. Such an observation is in line with the general principle of Chapter 2 that information for a registration task scales with the square of the signal gradient. Recognizing that regularization terms and smoothness constraints on the displacement field may violate the assumption of an unbiased estimator for deformable registration, the model described above nonetheless provides a useful basis of comparison for evaluation of registration performance — particularly in relation to image quality factors such as noise (dose) and spatial resolution.

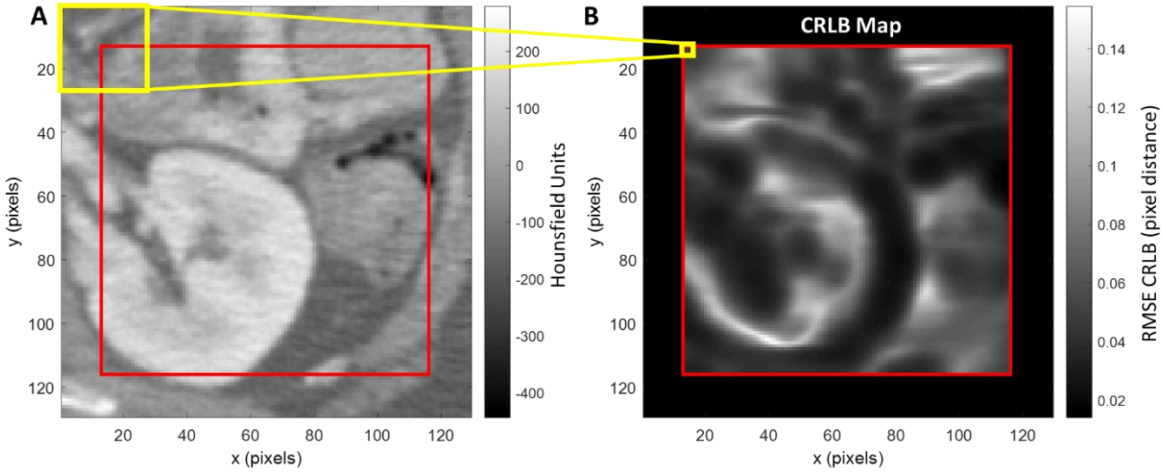


Figure 5.1: The CRLB "map" is formed by computing Eq. (2.12) over local image patches. (A) Example soft-tissue image patch. (B) CRLB map corresponding to the image in (A) where the contribution of an exemplary patch in (A) (yellow box) is related to the corresponding pixel in (B). Note the reduced CRLB about regions of high gradient. Figure adapted with permission of the publisher from [121].

5.2.3 Image Synthesis

Much of the work of the previous chapters has been centered on intra-modality registration, where both the preoperative and intraoperative image (in the scenario of image guidance) were from CT, CBCT, or x-ray projection images. A common scenario, however (e.g., image-guided neurosurgery), is that of having a preoperative MR scan that must be registered to an intraoperative CBCT scan. As there is no linear intensity relation between these two images, we can no longer apply the statistical model presented in Sec. 5.2.2, as there is no consistent true signal (g) term between the two images. Recent developments in image synthesis, however, allow us to frame the multi-modality registration as an intra-modality registration by first synthesizing one of the images to match the modality of the other.

Image synthesis is the process of translating (or "synthesizing") an image from one modality to another — for example, creating a synthetic CT image from an MR image. Generally, image synthesis methods are categorized according to whether the two modalities have paired or unpaired data. In the case of paired data — meaning each image in one of the modalities has a paired and geometrically aligned image in the other modality — an image-to-image [127] training method can be implemented, evaluating the loss directly between the synthesized image and the available ground truth. Paired data methods can rely on adversarial losses (by training a discriminator network), but can also incorporate image difference losses (e.g., L1 or perceptual losses) as is done in the Super Resolution GAN (SRGAN) [128].

Unpaired methods are utilized when there is no pixel-to-pixel correspondence established between the images in the two modalities — e.g., a set of CT and T1 MR volumes that are not aligned and may or may not have overlap in the human subjects that each contains. For this scenario, methods generally rely on frameworks similar to the Cycle-GAN [129], where the generators for both directions (e.g., MR-to-CT and CT-to-MR networks) are trained simultaneously. The general training workflow involves computing an adversarial loss on the synthesized images to ensure that they have general appearance of the target modality, and then re-synthesizing these images using the reverse-direction generator to compute an L1 loss between the input image and the re-synthesized image, thus acting as a regularizer to ensure spatial content is preserved in the synthetic image. The L1 loss is often not enough to fully preserve anatomical content in the synthesized image, and variations on the Cycle-GAN have introduced a structural constraint — e.g., MIND similarity [130] or Gradient Consistency [131] — between the input and synthesized images to preserve anatomical structure.

5.3 Experimental Methods

5.3.1 Deformable Registration Network Architecture

In this chapter, we use a supervised CNN approach in which ground truth deformations were simulated, ensuring that the errors observed arise from the network and training conditions, rather than potential errors in ground truth definitions or sub-optimality of the similarity metric. Based on the popularity of the UNet [132] and the potential for the results to be generalized to other CNN applications, this work examined a modified 2D version of the SVF-Net [123] for deformable image registration, which is based upon the UNet architecture. As illustrated in Fig. 5.2, the network takes the stacked 2D images (moving and fixed) to be registered as input and produces the displacement vector field as the output. The network was implemented in TensorFlow and was trained in a supervised manner with the ground truth displacement fields (discussed in Sec. 5.3.2) and optimized over an L2 loss function on the error in the predicted displacement field using the Adam optimizer [133] with a learning rate of 5×10^{-4}.

For comparison, we also examined the performance of conventional registration methods that are based on physical models (compared to learning-based methods). These included the Fast Symmetric Forces Demons algorithm [134] and B-spline FFD [115] as implemented in SimpleITK [63]. For each algorithm, we utilized a morphological pyramid and optimized the displacement field smoothing parameters (Demons) and number of control points (B-spline).

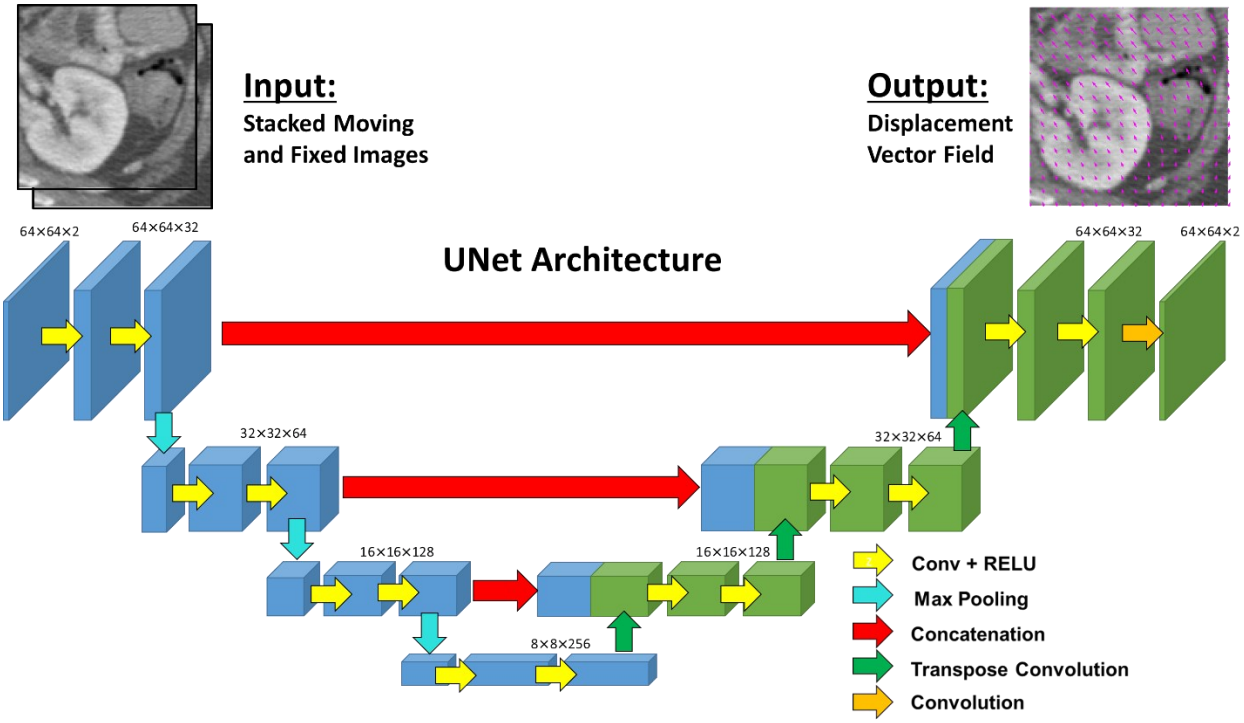


Figure 5.2: Convolutional neural network architecture adapted from SVF-Net for 2D (slice) image registration. The 2 stacked 64×64 image patches are supplied as input, and the output is the 2D 64×64 displacement vector field. Blue and green coloring of the features is included for improved visualization of the concatenation step. Figure adapted with permission of the publisher from [122].

5.3.2 Test Image Generation

Training and test images were generated by sampling from a Voronoi image distribution (Sec. 3.2), where seed points were uniformly and randomly sampled within the image, and piece-wise constant regions were subsequently defined by randomly sampling in the soft-tissue range of −110 to 90 Hounsfield Units (HU). Since piece-wise constant content produces degenerate solutions in deformable image registration, a small amount of clutter content (10 HU standard deviation) was added to the image by directly sampling from a $1/f^3$ distribution (matching that of the 2D Voronoi power spectrum). The image content was

cropped to a 32 cm diameter cylinder with isotropic 0.68 mm pixel size, yielding images as shown in Fig. 5.3A.

Ground truth displacement fields were simulated by sampling x- and y-components from a power-law ($1/f^{4.5}$) distribution to generate smoothly varying deformation. The displacement fields were applied to the noiseless images, after which realistic CT noise was injected into both the original and warped images. The noise injection process involved converting the image from HU to attenuation coefficients, performing 360 digital forward projections over 360°, injecting Poisson noise in the projection domain, and reconstructing using FBP. The magnitude of quantum noise was adjusted by scaling the fluence associated with the forward projection according to a specified dose level (quantified by tube-current time product, mAs) using the SPEKTR toolkit [61]. Furthermore, the spatial resolution in the image was adjusted by varying the cutoff frequency of the Hann apodization filter applied during reconstruction. The primary factor governing spatial resolution in this simulation was the apodization filter, and the FWHM of the point spread function was approximated as the inverse of the Hann cutoff frequency. Following noise injection, corresponding 64×64 pixel patches were sampled from the original and warped images for use as training and test data. While training minimized error on the full displacement field, evaluation on test data was performed by measuring the mean TRE at corner points within the test image patches (defined unambiguously by the intersection points among three Voronoi regions).

The process described above presented three distinct experimental parameters for investigating the effect of statistical mismatch between training and test data: (1) the image noise (i.e., quantum noise), controlled by variation of dose (referred to as $\mathrm{D_{train}}$ and $\mathrm{D_{test}}$); (2) the spatial resolution, controlled by variation of the FWHM (denoted $\mathrm{FWHM_{train}}$,

$\text{FWHM}_{\text{test}}$); and (3) the mean deformation magnitude (denoted $\overline{\text{X}}_{\text{train}}$, $\overline{\text{X}}_{\text{test}}$). Variations in the images associated with variation of these parameters are depicted in Fig. 5.3.

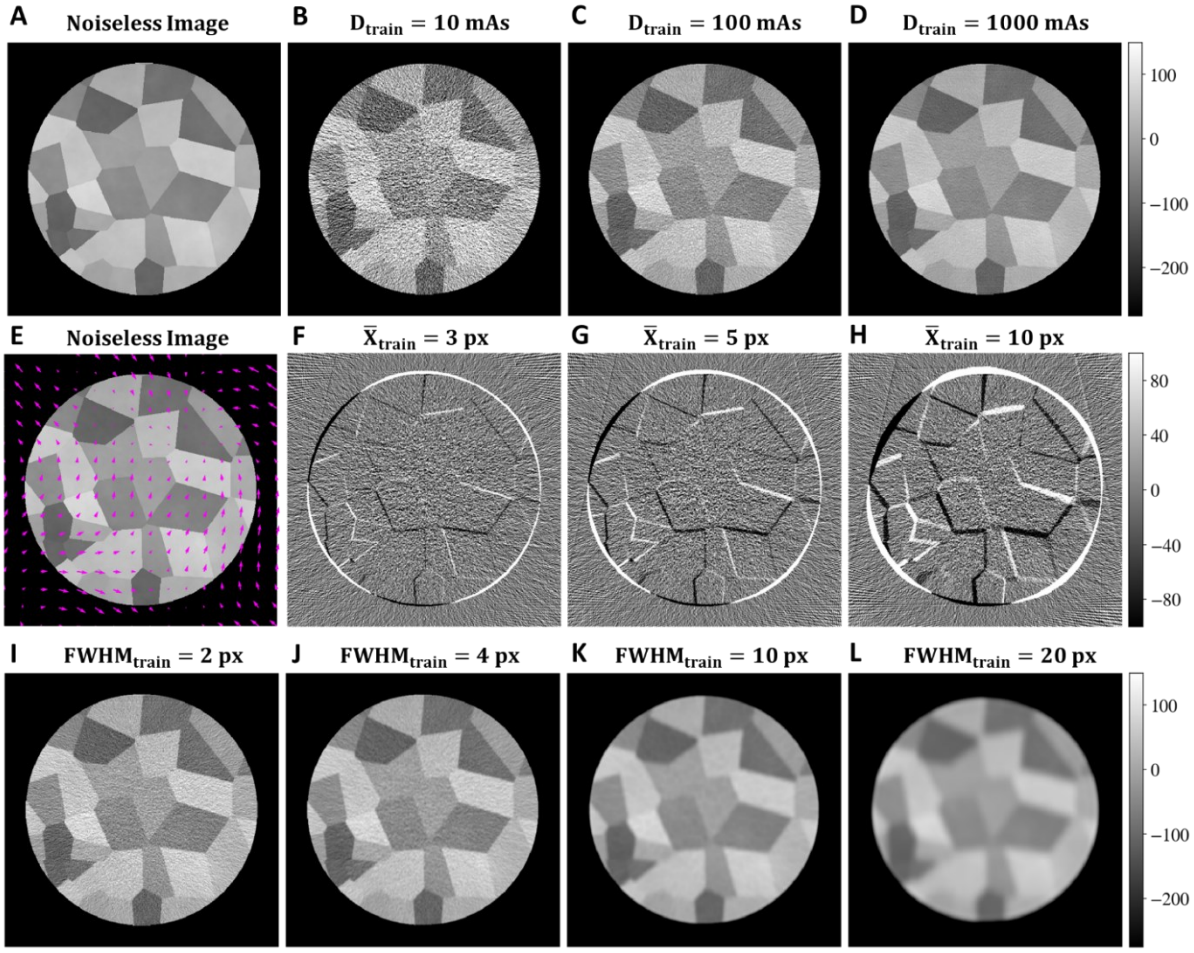


Figure 5.3: Image generation. The simulated noiseless image (A) is injected with noise to form the moving image with (B–D) showing example images at 3 dose levels (where dose is linearly related to the x-ray tube current-time product, mAs). Displacement vector fields are applied to the noiseless image (E) prior to noise injection to generate the fixed image with (F–H) showing the difference images of the fixed and moving images prior to registration for 3 levels of deformation magnitude. Variations on the apodization filter cutoff allows for reconstruction at various spatial resolutions (I–L). Figure adapted with permission of the publisher from [122].

5.3.3 Mismatch in Noise Magnitude

Training data in medical imaging, particularly in retrospective studies, are often limited in the diversity of dose levels exhibited. As a result, the dose levels observed during network

deployment could vary widely from those observed during training. The following experiments examined the effect of statistical mismatch of noise between training and test data.

(1) Single Dose Training: CNNs were trained with data from a single dose level (e.g., $D_{train} = 50$ mAs) on approximately 108,000 image patch pairs, each with $FWHM_{train} = 2$ px and $\bar{X}_{train}$ uniformly sampled from $[0.01, 0.1, 1, 3, 5, 10]$ px. For each dose condition (ranging from 5 to 1500 mAs), 11 networks were trained from random initialization, and the TRE was examined as a function of the difference of the dose in test data (D_{test}) from that in training data.

(2) Diverse Dose Training: Additional experiments examined the effect of training on a dataset containing a diverse range of dose levels. CNNs were trained on ~108,000 image patch pairs with dose levels uniformly sampled from [5, 10, 50, 100, 500, 1000, 1500] mAs. Additionally, a separate network was trained in a sparse manner, observing only two dose levels: 54,000 image patch pairs at 10 mAs and 54,000 image patch pairs at 1500 mAs. The networks were then evaluated by examining TRE as a function of the dose of the test image.

5.3.4 Mismatch in Image Resolution

Spatial resolution is another factor that is often variable in the population that could be sparsely represented in a training data set (e.g., a data set with all images acquired with the same make / model / manufacturer of scanner with particular post-processing / reconstruction protocols). The following experiments examined the effect of statistical mismatch in spatial resolution between training and test data.

(1) Single Resolution Training: CNNs were trained observing data from a single resolution level (e.g., $\mathrm{FWHM}_{\mathrm{train}} = 2$ px) on approximately 108,000 image patch pairs, each with $\mathrm{D}_{\mathrm{train}}$ uniformly sampled from [5, 10, 50, 100, 500, 1000, 1500] mAs and $\overline{\mathrm{X}}_{\mathrm{train}}$ uniformly sampled from $[\,1, 3, 5, 10]$ px. For each FWHM condition (ranging from 2 to 20 px), a single network was trained from random initialization, and the TRE was examined as the resolution of the test data ($\mathrm{FWHM}_{\mathrm{test}}$) diverged from that of the training data.

(2) Diverse Resolution Training: Additional experiments examined the effect of training on a data set containing a diverse range of resolution levels. CNNs were trained on ~108,000 image patch pairs with resolution levels uniformly sampled from Hann frequency cutoffs ranging from 0.1 to $1.0 \times f_{Nyq}$ (with increments of $0.1\times f_{Nyq}$), yielding $\mathrm{FWHM}_{\mathrm{train}}$ values ranging from 2 to 20 px. The network was then evaluated by examining TRE as a function of the resolution of the test image.

5.3.5 Mismatch in Deformation Magnitude

The magnitude and range of soft-tissue deformation is a statistical characteristic that is often difficult to control when curating a training data set and is perhaps even more difficult to control when the network is deployed in a particular application. Therefore, it is important to understand how the network behaves as the statistics of the deformation differ between the test and training data.

(1) Single Deformation Magnitude Training: CNNs were trained using data from only a single mean deformation magnitude level (e.g., $\overline{\mathrm{X}}_{\mathrm{train}} = 5$ px) on ~108,000 image patch pairs, each with $\mathrm{FWHM}_{\mathrm{train}} = 2$ px and $\mathrm{D}_{\mathrm{train}}$ uniformly sampled from [5, 10, 50, 100, 500, 1000,

1500] mAs. For each $\overline{X}_{train}$ condition (ranging from 0.01 to 10 px), a single network was trained from random initialization, and the TRE was examined as the mean deformation magnitude of the test data ($\overline{X}_{test}$) diverged from that of the training data.

(2) Diverse Deformation Magnitude Training: Additional experiments examined the effect of training on a dataset containing a diverse range of mean deformation magnitude. CNNs were trained on approximately 108,000 image patch pairs with $\overline{X}_{train}$ uniformly sampled from $[0.01, 0.1, 1, 3, 5, 10]$ px. The network was then evaluated by examining TRE as a function of the mean deformation magnitude of the test image.

5.3.6 Testing on Anatomical Image Content

Networks trained on Voronoi images in the above experiments were applied to registration of real anatomy (a patient image from an IRB-approved study) in axial CT abdominal images (proximal to the kidney, as in Fig. 5.4A). The noise injection and deformation process described in section 2.3.1 was applied to the 128×128 pixel image to generate a registration scenario with $D_{test} = 500$ mAs, $FWHM_{test} = 2$ px, and $\overline{X}_{test} = 3$ px, yielding a fixed and moving image. The difference image prior to registration is shown in Fig. 5.4B. Note that due to the architecture, even though the network was trained on 64×64 patches, it can be deployed on the 128×128 px image (and larger power of 2 image sizes) without modification — with some performance differences arising from changes in boundary conditions.

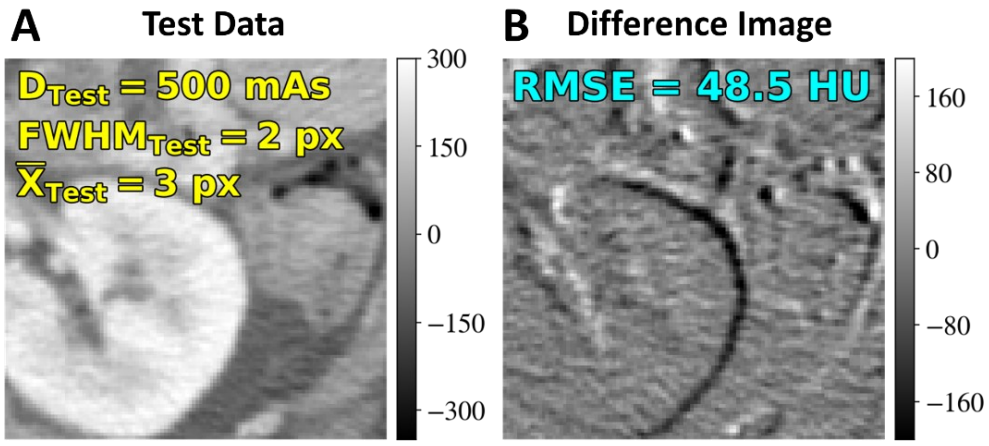


Figure 5.4: Testing on anatomical content after training on Voronoi images. Moving (A) and fixed images were generated at $D_{test} = 500$ mAs, $FWHM_{test} = 2$ px, and $\bar{X}_{test} = 3$ px, yielding the difference image in (B). Figure adapted with permission of the publisher from [122].

The noise injection/deformation process was similarly applied to 10 abdominal CT images from The Cancer Imaging Archive (TCIA) [135] to examine the performance of the diversely trained networks on real anatomy. Each of the 10 images were reconstructed under various conditions of dose, resolution, and deformation magnitude, and cropped to 128×128 pixel image patch pairs focusing on soft-tissue regions of interest. For each experimental condition, the noise and deformation injection process was repeated 10 times per image, yielding 100 total image pairs. In each image, 10 conspicuous soft-tissue anatomical landmarks were selected for evaluation of TRE. The diversely trained networks were then deployed on these image patch pairs, and TRE was examined as a function of test image dose, resolution, and deformation magnitude.

5.4 Results

5.4.1 Registration Results: Effect of Noise Mismatch

Figure 5.5 shows TRE performance as a function of test image dose (with $\overline{X}_{test} = 3$ px, $FWHM_{test} = 2$ px) for the conventional registration methods and the CNN-based method at several training conditions. To provide context, the results were assessed relative to the bounds imparted by three figures of merit: (1) the "Initial Error" line, referring to the error associated with predicting a null displacement field; (2) the D_{train}, D_{test} = "Noiseless" line, referring to the performance when train and test data are noiseless, yielding an optimal bound that noisy data should not exceed; and (3) the CRLB for rigid registration, indicating ideal registration performance as a function of dose for unbiased estimators.

Figure 5.5A illustrates CNN registration performance in the statistically matched case ($D_{train} = D_{test}$, red line) where the dose of the training data exactly matches that of test data; each point shows the TRE (mean ± std) for 11 networks trained at that dose level with 11 random initializations (e.g., the data point at $D_{test} = 100$ mAs indicates the performance of networks trained at $D_{train} = 100$ mAs). The $D_{train} = 10$ mAs and $D_{train} = 5$ mAs datapoints only show the results of 3 and 2 trained networks, respectively, as most of the 11 randomly initialized networks did not successfully converge under these conditions, indicating the high sensitivity associated with training only on very noisy data. Generally, we see that CNN registration error was reduced with higher dose and yielded comparable or better performance to the conventional registration methods — outperforming the conventional methods in the low-dose range and achieving sub-pixel TRE in the high-dose range (down to 0.5 px TRE at

1500 mAs). The Demons and CNN methods appeared to trend similarly as a function of dose, whereas B-Spline FFD presented a steeper reduction in error with increased dose. None of the methods, however, closely followed the $\sim 1/\sqrt{\text{dose}}$ trend set by the CRLB, indicating that the assumption of independent, locally rigid registration is a weak approximation to deformable registration, although the CRLB still appears to present a reasonable lower limit to performance.

Figure 5.5B shows CNN registration performance for networks trained with only a single dose level. Examination of the high-dose training $\mathrm{D_{train}} = 1500\ \mathrm{mAs}$ condition shows similar performance to the $\mathrm{D_{train}} = \mathrm{D_{test}}$ case for a large range of test image dose levels — down to approximately 50 mAs, where the registration performance begins to diverge from the statistically matched condition. Interestingly, for the CNNs trained at lower dose levels, we observe that registration performance plateaus (and even slightly increases) as the dose of the test image exceeds that of training images, indicating that there is no benefit in deploying the network on images acquired at higher dose (and lower noise) than the training data. Furthermore, these lower-dose training conditions did not exhibit a large range of robustness, where networks trained at 50 mAs yielded similar performance to those trained at 1500 mAs when tested on 10 mAs images (2.73 vs. 2.52 px mean TRE).

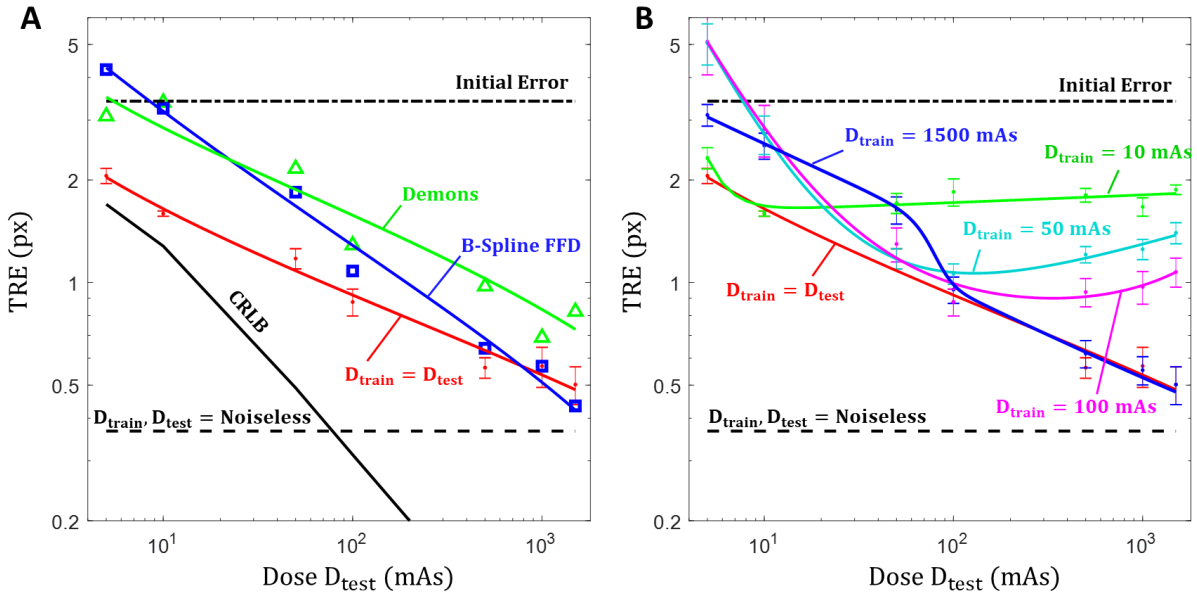


Figure 5.5: Registration performance as a function of test image dose. (A) TRE as a function of D_{test} for single-dose training statistically matched CNN ($D_{train} = D_{test}$, red), Demons (green triangle), and B-Spline FFD (blue square). These results are generally bounded by the rigid CRLB (black line), the Initial Error line (black dot-dash), and the D_{train}, D_{test} = Noiseless error (black dashed). (B) TRE as a function of D_{test} for single-dose training CNNs showing the effect of mismatched statistics for D_{train} values of 10 (green), 50 (cyan), 100 (magenta), and 1500 (blue) mAs. Figure adapted with permission of the publisher from [122].

Dashed curves in Fig. 5.6 show the registration error as a function of test image dose for CNNs trained at diverse dose conditions. First, we observe that performance for the single network trained on a diversity of images with dose levels ranging from 5–1500 mAs closely matched the performance of the multitude of networks associated with the $D_{train} = D_{test}$ curve, with only a slight reduction in performance in the very high dose region. Furthermore, the network trained at only two dose levels, with half the training data at 1500 mAs and half at 10 mAs, yielded nearly the same performance as the highly diverse D_{train} = 5–1500 mAs network, indicating that a wide *range* of dose levels (not necessarily densely or uniformly sampled) may be sufficient to diversify the training set.

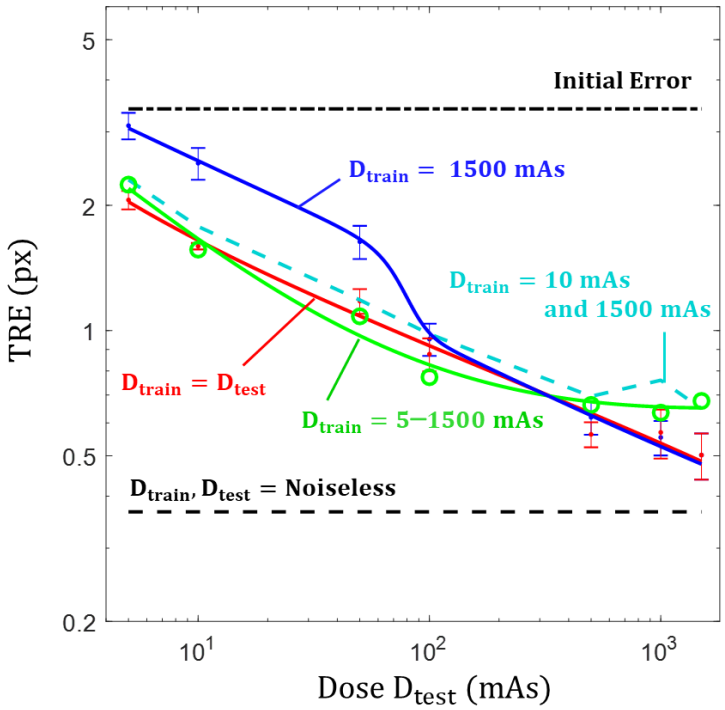


Figure 5.6: Diverse dose training. The green (D_{train} = 5–1500 mAs) line shows TRE performance for the diversely trained (with respect to dose) network and the cyan dashed line depicts error when half the training data was 10 mAs and half was 1500 mAs. The blue (D_{train} = 1500 mAs) and red (D_{train} = D_{test}) solid lines from Fig. 5.5 are provided for reference. Figure adapted with permission of the publisher from [122].

5.4.2 Registration Results: Effect of Image Resolution Mismatch

Figure 5.7 shows the TRE measured as a function of the spatial resolution (FWHM) in the test images (with $\overline{X}_{test}$ = 3 px, D_{test} = 1500 mAs). The $FWHM_{train}$ = 2 px curve (magenta) shows the performance of a network trained on high resolution images, where we observe a linear increase in error as networks are tested on lower resolution images. The $FWHM_{train}$ = 4 px training (cyan) provides increased robustness (compared to $FWHM_{train}$ = 2 px) in the low-resolution test region, and performance is only slightly reduced in the high-resolution range. However, training on very low-resolution data ($FWHM_{train}$ = 10 px [red] and $FWHM_{train}$ = 20 px [blue]) does not generalize to high-resolution test data, with a steep

increase in error as the resolution of the test data exceeds that of the training data. We see again the diverse training network (green) generalizes well, providing near optimal performance across the entire range of tested image resolution levels. Comparison of the CNN performance with the conventional methods initially indicates that the conventional methods nearly always outperform the network, however this can be attributed to two factors: (1) each data point for the conventional methods represents the TRE for best performing parameter selection at that FWHM test condition therefore it represents a "best-case" for the conventional methods; and (2) the testing is performed on high-dose images where similar performance was observed (Fig. 5.5A) among the conventional and CNN-based methods.

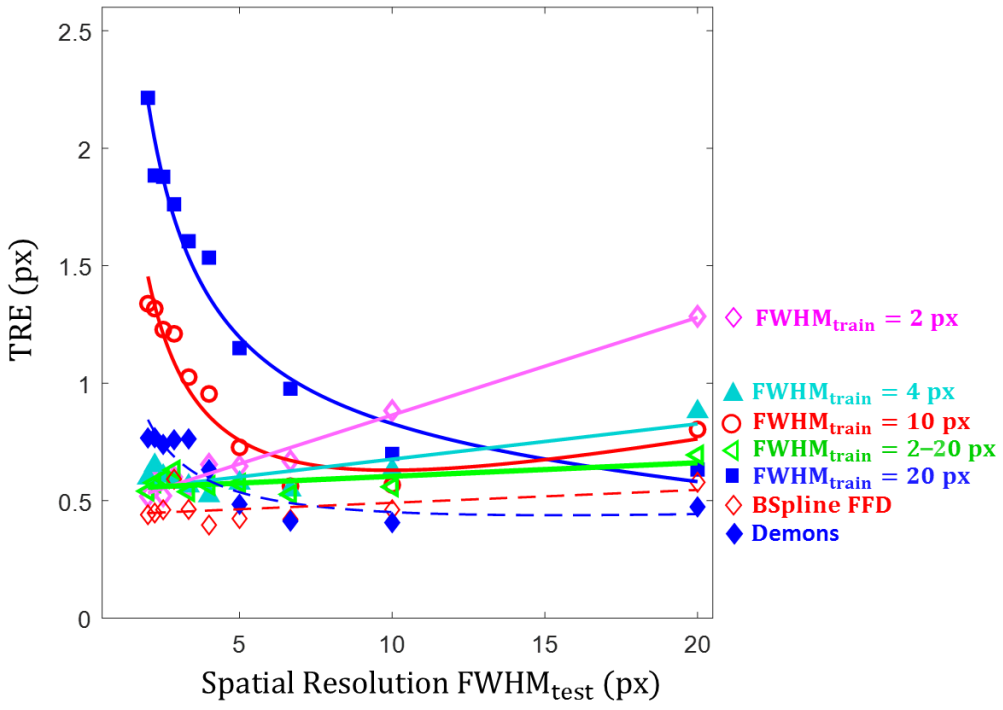


Figure 5.7: Effect of image spatial resolution. TRE results as a function of $FWHM_{test}$ for CNNs trained at various spatial resolutions: $FWHM_{train} = 2$ (magenta diamond), 4 (cyan triangle), 10 (red circle), and 20 (blue square) px. The green line ($FWHM_{train} = 2–20$ px, sideways triangle) shows registration performance for the diversely trained (with respect to FWHM) network. Dashed lines show the performance of Demons and B-Spline FFD for comparison. Figure adapted with permission of the publisher from [122].

5.4.3 Registration Results: Effect of Deformation Mismatch

Figure 5.8 shows registration performance of the CNN as a function of $\overline{X}_{test}$ (with $FWHM_{test} = 2$ px, $D_{test} = 1500$ mAs) for networks trained with fixed mean displacement magnitude, $\overline{X}_{train}$. The experimental error generally increases as the deformation magnitude increases (as the registration becomes more difficult to solve); however, among each $\overline{X}_{test}$ deformation level, the best performance is observed when training and test data are statistically matched (i.e., $\overline{X}_{test} = \overline{X}_{train}$). At large deformation ($\overline{X}_{test} = 10$ px), subpixel error is no longer achieved, and the best performing network at that condition ($\overline{X}_{train} = 10$ px) exhibited a mean TRE of 1.68 px. While networks generalized well when $\overline{X}_{test} < \overline{X}_{train}$, a sharp increase in error occurs if $\overline{X}_{test}$ exceeded the mean displacement magnitude of the training conditions. However, it should be noted that high $\overline{X}_{train}$ data will still likely contain regions of small deformation, aiding the ability to generalize. The diversely trained $\overline{X}_{train} = 0.01$–$10$ px condition yielded a good compromise on performance across all test conditions.

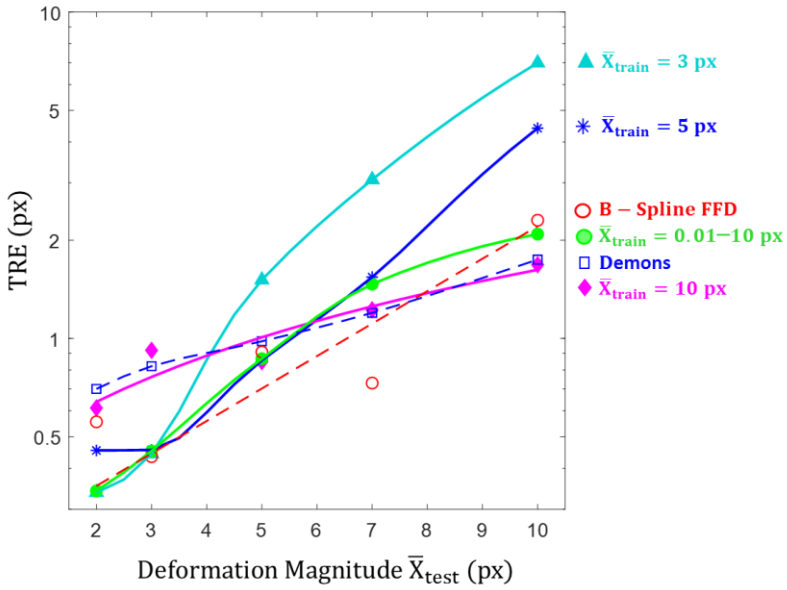


Figure 5.8: Effect of mismatch in mean deformation magnitude. TRE measured as a function of mean displacement magnitude for CNNs trained at $\overline{X}_{train} = 3$ (cyan triangle), 5 (blue star), and 10 (magenta diamond) px. The green line $\overline{X}_{train} = 0.01–10$ px, circle) shows registration performance for the diversely trained (with respect to $\overline{X}$) network. Dashed lines show the performance of Demons and B-Spline FFD for comparison. Figure adapted with permission of the publisher from [122].

5.4.4 Registration Results: Testing on Anatomical Image Content

Figure 5.9 shows the registration performance for networks trained on Voronoi content alone and tested a real anatomy in axial CT abdominal images. Registration results are shown in terms of the difference images following registration, with the RMSE difference in pixel intensity (HU) shown in each case. The rows are organized according to the three prior experiments, examining the effect of dose, resolution, and deformation magnitude, respectively. The columns represent three conditions: mismatched statistics, matched statistics, and diverse training. Considering the difference images and RMSE values following registration, we observe results consistent with the results described above — namely, that

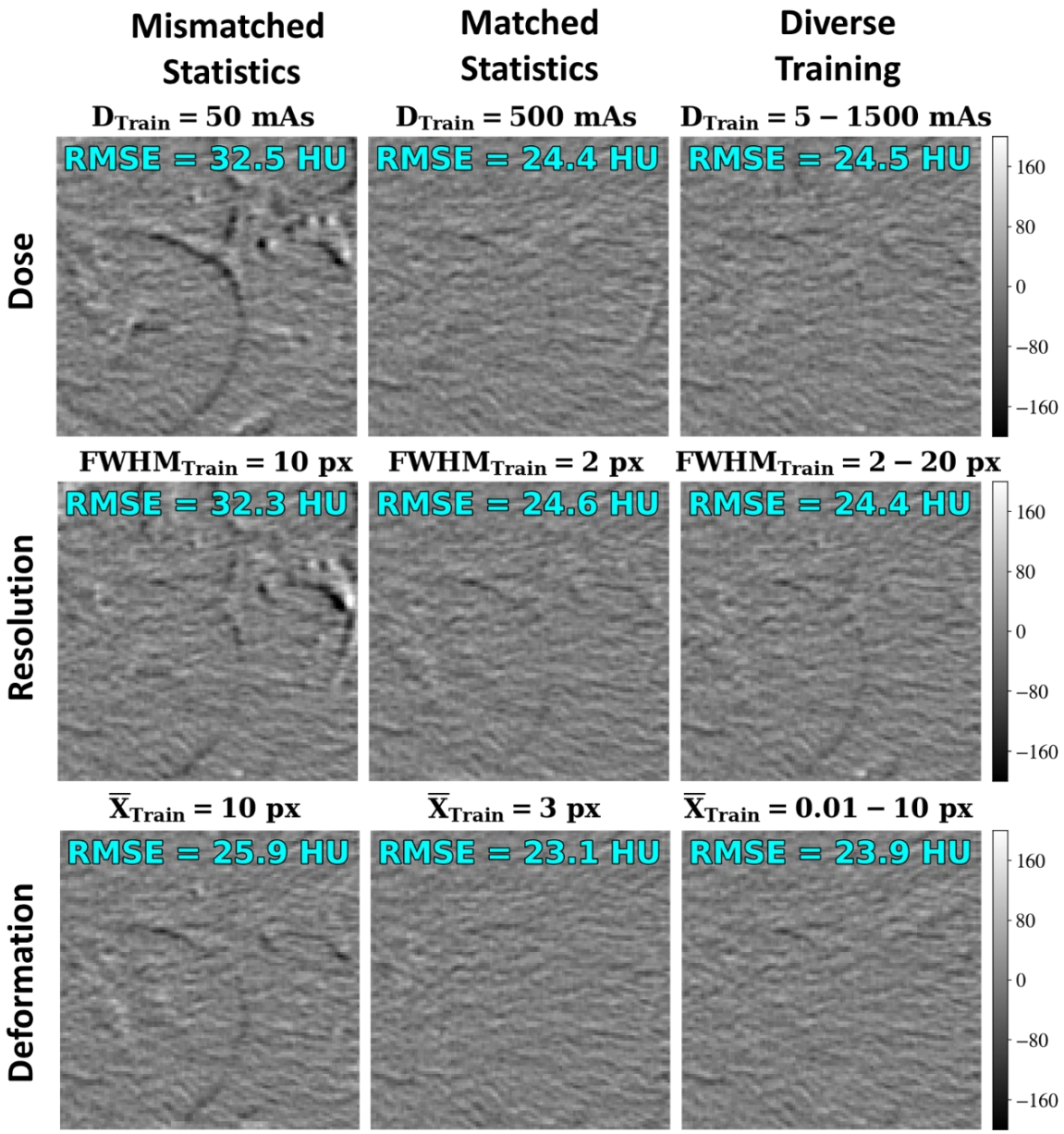


Figure 5.9: Testing on anatomical content. Difference images following registration (original images shown in Fig. 5.4) are shown for networks at various training conditions. RMSE of the difference in HU shown in text for each image. Columns represent conditions of either mismatched training and test statistics, matched statistics, and diverse training. Rows examine various training conditions for dose, resolution, and deformation magnitude. Figure adapted with permission of the publisher from [122].

matching the statistics of the training data to those of the test data tends to be optimal, but training on diverse datasets provides comparable (and generally more robust) performance. Furthermore, it is promising that training on Voronoi images alone yielded reasonable registration performance in real anatomy, providing another validation to the Voronoi training model.

Figure 5.10 further demonstrates the performance of the diversely trained networks applied to images of real anatomy. Figure 5.10 shows the distributions (mean ± 1 standard deviation, computed over 100 image pairs) of the mean TRE for each image pair (mTRE, computed from 10 landmarks per image pair) for the diversely trained networks. Figure 5.10A shows the performance for the diversely trained network (D_{train} = 5–1500 mAs) and demonstrates a reduction in mTRE with increased dose (holding $\overline{X}_{test} = 3$ px and $FWHM_{test} = 2$ px). The mean of the mTRE measurements exhibits a 1/sqrt dependence on dose (R^2 = 0.98) in agreement with the statistical model presented in Sec. 2.2.2.2. Example images representing the median performance at low and high dose levels are shown below each plot, with Canny edges overlaid on the registered image. Similarly, Fig. 5.10B shows the performance of the diversely trained ($FWHM_{train}$ = 2–20 px) network applied to images generated at various levels of spatial resolution (holding $D_{test} = 1500$ mAs and $\overline{X}_{test} = 3$ px). A nonmonotonic (quadratic) dependence on spatial resolution ($FWHM_{test}$) is exhibited (R^2 = 0.91) with weak correlation to $FWHM_{test}$ (~0.1 px variation in mean mTRE over the full range of $FWHM_{test}$). Finally, Fig. 5.10C shows results for the diversely trained ($\overline{X}_{train}$ = 0.01–10 px) network as a function of the test image deformation magnitude (holding $FWHM_{test} = 2$ px and $D_{test} = 1500$ mAs), also demonstrating a roughly quadratic dependence (R^2 = 0.99) on the mean fit. While the trends in mean of the mTRE measurements

are consistent with basic models of performance (e.g., 1/sqrt dependence on dose), the individual mTRE measurements exhibit high variability, and fitting the collection of mTRE measurements (rather than the per-condition mean) to the models tested above exhibits low correlation ($R^2 = 0.11,\ 0.02, \text{and } 0.43$ for Figs. 5.10A–C, respectively). Thus, the experimental variables are not strongly predictive of mTRE for a single image (e.g., a noisy image may spuriously yield more accurate registration than a higher dose image); however, the overall trends in mean mTRE were as expected. Overall, image registration is robustly achieved, except perhaps for the case of large test image deformation in Fig. 5.10C. In each case, the trends in TRE reflect that of the registration errors shown above for the diversely trained networks applied to Voronoi test images (Figs. 5.6, 5.7, and 5.8, respectively), again validating the use of Voronoi content as a statistical model for registration training that appears to transfer reasonably well to registration of real anatomy.

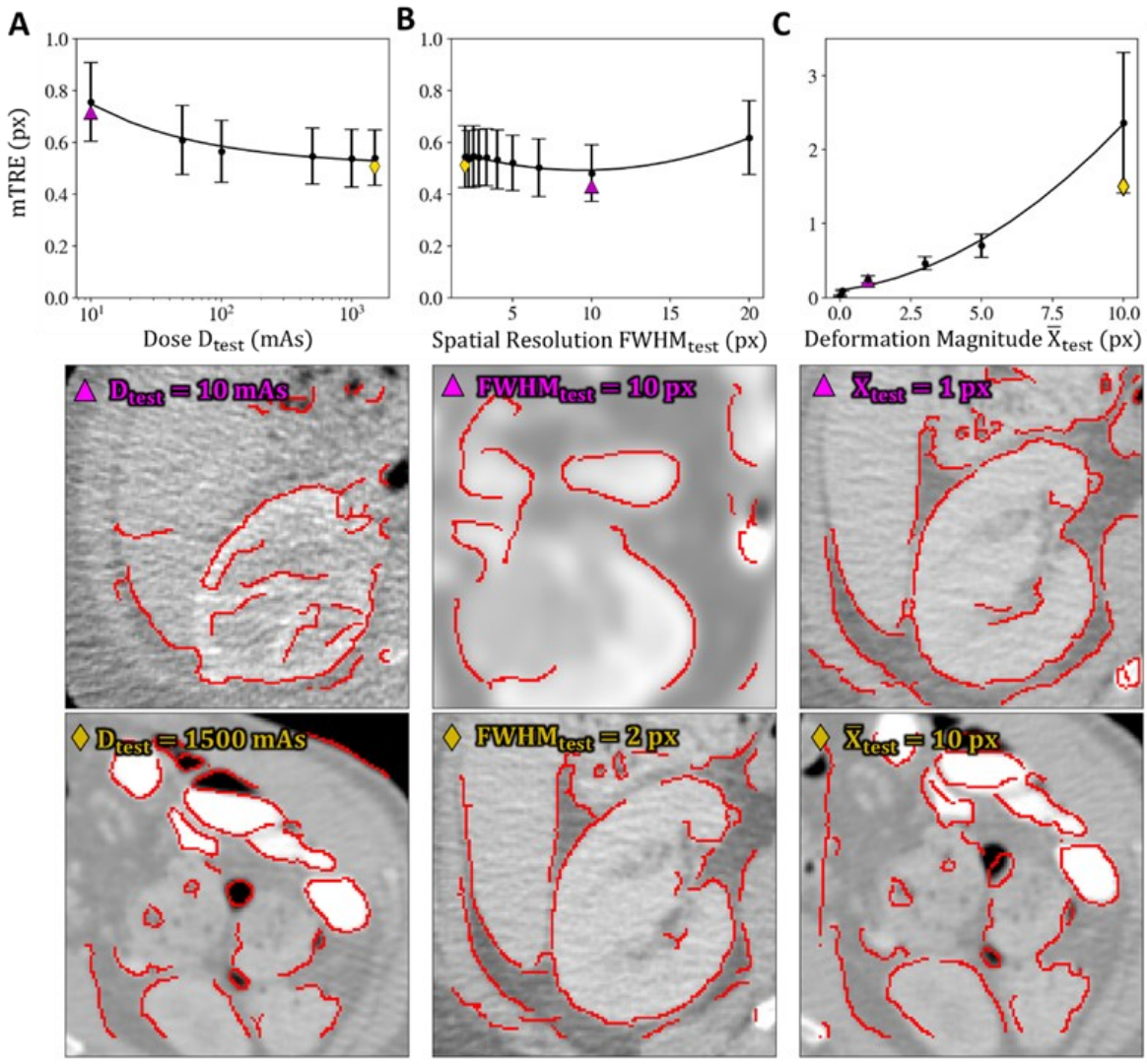


Figure 5.10: Registration error mTRE (mean ± 1 standard deviation) of the diversely trained networks applied to anatomical content as a function of the TCIA test image (A) dose, (B) spatial resolution, and (C) deformation magnitude. Below each plot, are the median performers for two test conditions with the symbol on the image referring to the plotted symbol in the associated graph. Figure adapted with permission of the publisher from [122].

5.5 Case Study: Image Synthesis for Multi-Modality Registration

5.5.1 Synthesis Method

We extend the understanding gained from previous sections investigating how CNN-based registration performance depends on the image statistics in the training data to application in other CNN methodologies — in this case image synthesis. Below, we examine the development of an MR-to-CT synthesis method used as a front-end to deformable registration of T1 MR and CBCT images (i.e., converting the moving MR image to a synthetic CT image for intra-modality registration to the fixed CT image). While typically multi-modality registration of this form would be achieved by optimizing over an inter-modality similarity metric (e.g., MI); by performing MR-CT synthesis we may utilize the insights gained from previous chapters for intra-modality registration, such as the importance of image gradient content and the optimality of CC-based metrics in the absence of content mismatch. Please note, however, that this section only investigates the CNN-based synthesis step.

As we observed in Sec. 5.4, statistical diversity in the training data was shown to be an important factor with respect to model generalization. A similar concern for model generalizability arises in the case of T1 MR-CT synthesis, where due to differences in scan protocol and vendors, wide variations in the slice thickness and anatomical feature contrast are present within the category of T1 MR images. For example, in this study the training data consists of 45 paired and geometrically aligned T1 MR and CT brain images, where each of the T1 images are high-quality gradient-echo scans with ≤ 1 mm voxel spacing (example slice depicted in Fig. 5.11A). However, much of the test dataset contains images from standard T1

acquisitions with 4–5 mm slice thickness and significantly reduced gray-white matter contrast as shown in Fig. 5.11B. In this study, we utilized our understanding of the importance of statistical diversity in the training data on network generalization in order to train an MR-CT synthesis network that generalizes well on standard T1 data, even though the network was trained primarily on gradient-echo data.

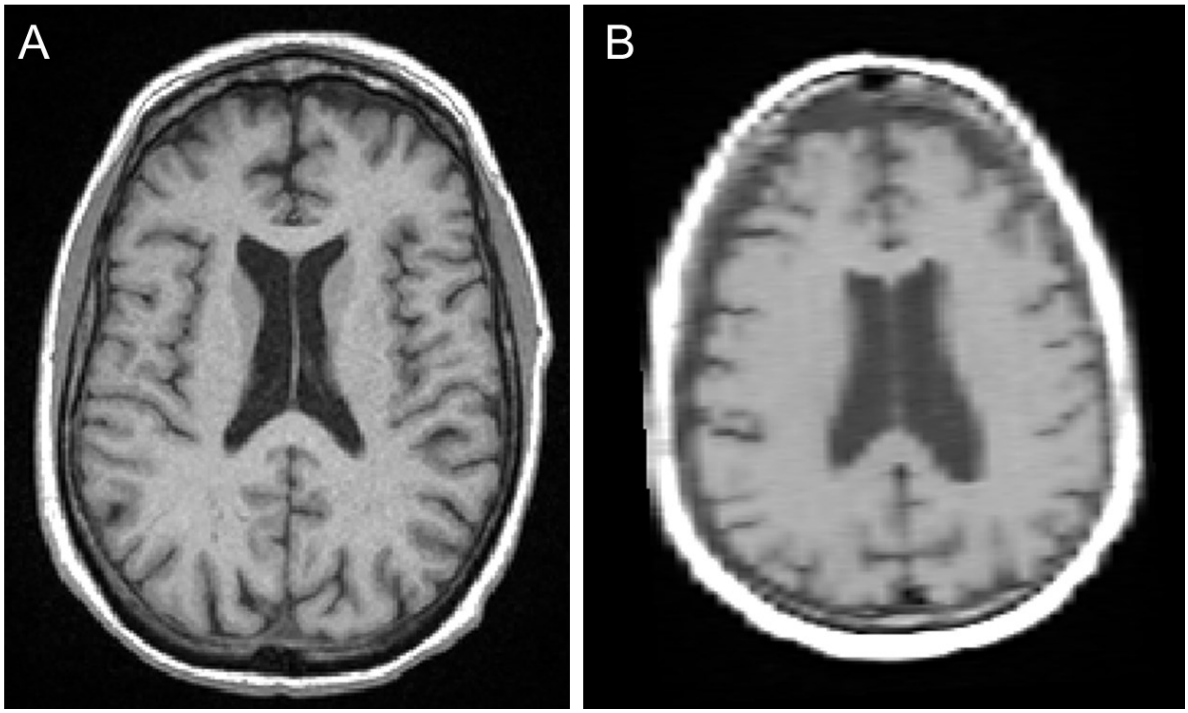


Figure 5.11: Example T1 MR image slices from the (A) gradient-echo T1 acquisition and (B) standard T1 acquisition.

We trained a 3D UNet to perform MR-CT synthesis using 45 high-quality T1 MR image volumes, where each volume was paired with a geometrically aligned CT volume. As the data was paired, we designed the UNet to directly learn the MR-CT synthesis (cf., an unpaired CycleGAN method) through a combination of an L1 loss between the true and synthesized CT and an adversarial loss on the synthesized CT. From the results of Sec. 5.4, we know that for the network to generalize well to standard T1 acquisitions, the statistical characteristics of the

standard T1 acquisition must be represented in the training data. In the study below, we explored three strategies for augmenting the gradient-echo training data to simulate standard T1 image appearance characteristics. Each subsequent strategy was be incorporated in addition to the previous strategy and more fully captured the appearance characteristics of standard T1 images. As this is an exemplary case study, many of the implementation details will be omitted for brevity so that we may maintain the connection to the core work.

Strategy 1: The first strategy is the simplest form of simulating standard T1 data, operating by augmenting the gradient-echo training data using a contrast level curve designed to reduce the gray-white matter contrast. A contrast level curve was hand-designed to reduce the contrast in the brain gray and white matter intensity range to mimic that of the standard T1 acquisition, and the curve was fitted to a high-order polynomial to allow fast augmentation of the image during training. Furthermore, from our observations, the non-cortical (spongy) bone tended to present with higher relative intensity in the standard acquisition than in the gradient-echo; therefore, a random multiplicative scaling was applied to this spongy bone region during augmentation. As the network still needs to perform well on gradient-echo test data, the augmentation strategy was not applied for approximately 10% of the training iterations.

Strategy 2: The second strategy more fully simulates the characteristics of the standard T1 acquisition by utilizing a gradient-echo-to-standard T1 image synthesis network. An unpaired 2D Disentangled Representation Network [136] was trained using the gradient-echo dataset and 7 standard T1 volumes. Similar to the CycleGAN, the Disentangled Representation Network learns a bi-directional synthesis using unpaired data; however, rather than directly performing image-to-image synthesis, it utilized an encoder-decoder framework where the encoder decomposes the input image into separate appearance and content representations. In

line with the work of [130], we further incorporated a structure-constrained loss between the input and synthesized image to ensure anatomical structure was preserved during synthesis. The gradient-echo-to-standard T1 network was then used to augment the training data for MR-CT synthesis training. The augmentation strategy was incorporated along with that of Strategy 1, where the augmented data was evenly distributed between the two strategies.

Strategy 3: The third strategy includes just one paired CT and standard T1 dataset along with the 45 gradient-echo datasets. The dataset was replicated within the training data so that it was observed during 10% of the training iterations. The previous augmentation strategies were still applied on the gradient-echo data during training.

Along with the above augmentation strategies, standard augmentation methods were applied in all cases. These augmentation strategies included random image rotations, jitter, linear scaling, thick slice simulation, and Rician noise injection.

5.5.2 Synthesis Results

Figure 5.12 shows example synthesis outputs for the three augmentations strategies (bottom row) applied to a standard T1 volume (top left). In the bottom left we see the output of Strategy 1, where only simple contrast level augmentation was performed. Compared to the reference CT (top right), we see an appreciable lack of contrast in the cerebral ventricles and a lack of definition for many of the sulci which are present in the reference CT.

Strategy 2 more fully incorporates the appearance characteristics of standard T1 images in the training data by using a synthesis network as an augmentation technique. In the resulting image we observe a small improvement in ventricle contrast (approximately 5–10 HU) and an

improvement in sulci definition in the anterior brain, though edge definition is still lacking among many of the structures.

Finally, we observe the best qualitative synthesis performance in Strategy 3 where clear definition of both the ventricles and sulci is present, though we note there are still some limitations in the spatial resolution due to the thick slice (5 mm) T1 input. The trend in qualitive improvement for each subsequent strategy is further reflected in the mean absolute errors (MAE, computed near the ventricles using the reference CT as ground truth), where we see Strategy 3 provided the lowest MAE.

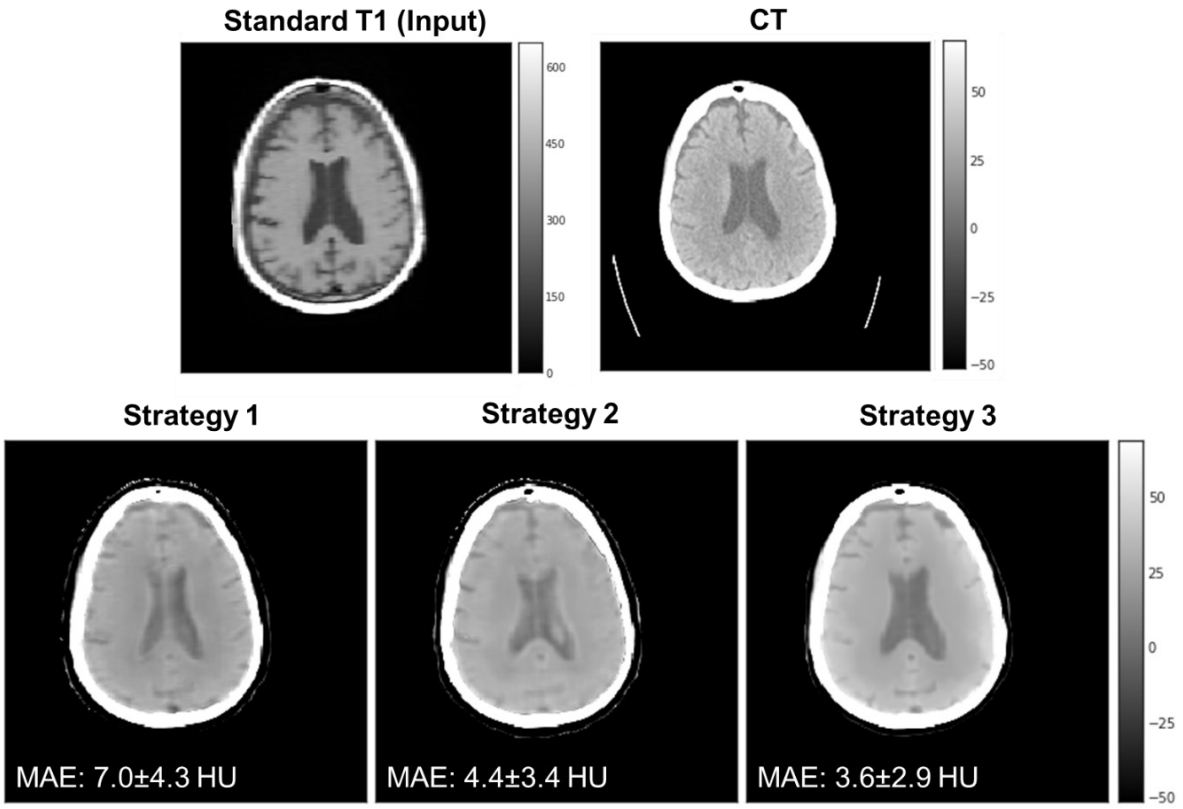


Figure 5.12: Synthesis outputs (bottom row) for the Standard T1 test data input (top left) using the three augmentation strategies. Patient CT (top right) provided as reference. MAE (mean±std) using the reference CT as truth is provided in the image for each strategy.

Figure 5.13 shows the lines profiles for Strategies 2 (green) and 3 (red) compared to the reference CT (black). From these profiles we observe both strategies yield a total contrast that matches that of the reference CT, however Strategy 2 contains overly blurred ventricle edges while Strategy 3 presents sharp ventricle edges that closely match the reference CT. Based on our model from Chapter 2 and the extension presented in Sec. 5.2.2, we know the information associated with a registration task is proportional to the square of the signal gradient content. Therefore, the degradation to the ventricle edges in Strategy 2 will only act to inhibit registration performance, making the Strategy 3 output better suited for a registration task.

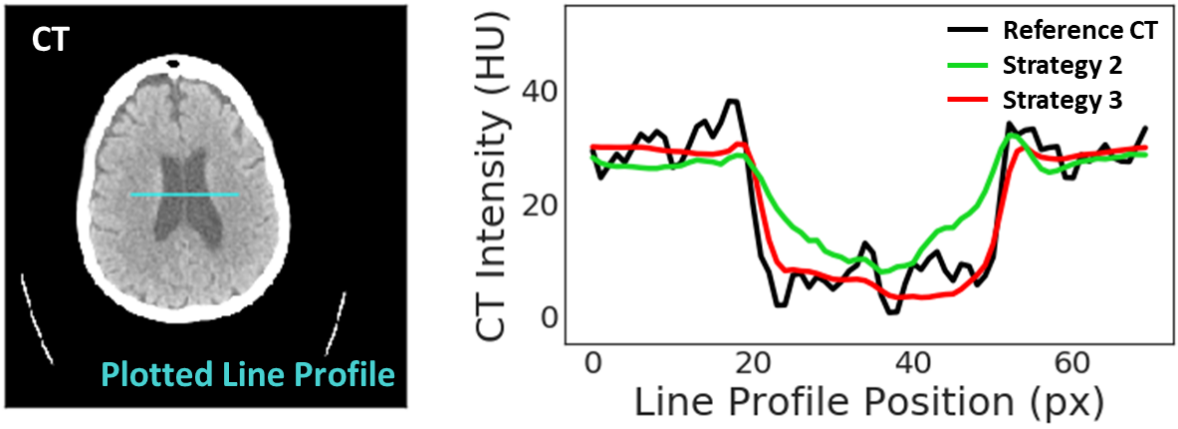


Figure 5.13: Plotted line profiles for the reference CT (black), synthetic CT from Strategy 2 (green), and synthetic CT from Strategy 3 (red).

5.6 Discussion and Conclusions

The quality of medical images — that is, the statistical characteristics underlying such images — varies widely, depending on the imaging system, image acquisition protocol (e.g., dose level), reconstruction method (e.g., smoothing filters), and post-processing techniques. Accordingly, it can be difficult to curate training data that is fully representative of the statistics to be encountered in testing and application in real clinical use. Therefore, an understanding of the behavior of the network as the statistics of the test data deviate from those of the training

data helps to ensure reliability of the network and/or determine whether additional data collection or augmentation is necessary. In this work, we specifically studied statistical mismatch in the form of image noise, spatial resolution, and deformation magnitude, finding generally that exactly matching the statistics is optimal; however, training the network with data featuring a diversity of statistical characteristics yields a single model that tends to be robust across a broader range of test conditions.

The experiments in this work provided insight on the importance of the various statistical characteristics that were examined. For mismatch in dose, it was found that testing on higher dose images than present in the training set did not improve (and in fact, slightly diminished) registration performance. Although testing on images that were noisier (e.g., lower dose) than the training data generalized well, the ability to maintain similar performance to the matched statistics case was limited. Performance was further improved by including a diverse range of dose levels in the training set, extending generalizability of the network, especially in the low-dose range (although slightly diminishing performance in the high-dose range). Interestingly, a training set composed of two distinct noise levels (e.g., high-dose data and low-dose data) yielded similar improvement, with performance comparable to that of a training set representing "all" intermediate noise levels — indicating that possibly just a wide *range* in statistical variation is necessary for generalization.

For mismatch in spatial resolution, while there was a modest reduction in performance by testing on blurrier data than present in the training set (which is easy to account with blurring augmentation methods), testing on much higher resolution images was found to exhibit a steep reduction in performance. Networks trained only on low-resolution data sets are therefore unlikely to extend well to high-resolution test data.

With respect to deformation magnitude, we found it important to ensure that the test data deformation magnitude did not exceed that observed in the training data. While the networks generalized better to smaller deformation scenarios, performance could be greatly improved by ensuring a wide range of deformation magnitude in the training data, which can be accomplished by augmenting the dataset with known deformations of various magnitudes.

In Sec. 5.5, we showed how diversity in the training data is also important in CNN-based methods for MR-CT synthesis when generalizing across the range of T1 MR acquisition protocols. The trend in performance for these augmentation strategies emphasizes the insight gained from Sec. 5.4 that a broad range of statistical variation must be included to achieve desirable network performance. With the addition of each strategy, we more fully captured the variation in appearance associated with standard T1 images, and the performance of the synthesis output was improved accordingly. Interestingly, the introduction of just a single standard T1 dataset in Strategy 3 yielded a significant improvement in ventricle edge definition compared to Strategy 2, indicating again that having the *range* of statistical variation represented (by at least one or a few samples in the training set) may be sufficient for generalization, and not necessarily a fully sampled uniform distribution.

Chapter 6: Summary and Conclusions

Thesis Statement: ***Statistical modeling of the factors that govern image registration accuracy provides a foundation for understanding the fundamental limits of registration accuracy and a quantitative guide to selecting imaging protocols and registration parameters that maximize the performance of image registration algorithms.***

6.1 Aim 1: Develop a Statistical Model Relating Image Quality to Image Registration Accuracy

In Chapter 2, two figures of merit were derived in the context of translation-only image registration, namely the CRLB and the CC RMSE estimate. Analogous to the prewhitening and non-prewhitening observer models that have been foundational in SDT for tasks of image detection, these figures of merit directly relate fundamental image quality metrics (i.e., NPS and MTF) and content characteristics (i.e., signal power spectrum) to the task of image registration. From a derivation of the relationship between the CRLB and such image quality factors, an understanding was immediately gained regarding the dependence of image registration accuracy on the performance of the imaging system (e.g., by relation to NEQ and

DQE) and post-processing parameters (e.g., smoothing width of Gaussian blur). The figures of merit were further validated under simulated registration experiments where the measured performance closely followed model predictions with respect to both dose and post-processing blur.

6.2 Aim 2: Develop a Statistical Model for Soft-Tissue Deformation in Rigid Image Registration

In Chapter 3, the framework of Aim 1 was extended to model the confounding role of soft-tissue deformation in rigid registration. In line with the field of SDT, which has modeled background anatomy as a confounding influence that can be treated as noise, theoretical and experimental analysis similarly showed that soft-tissue deformation can be modeled as a noise source in rigid image registration. Moreover, the effect can be quantified by incorporating the power spectrum of the soft-tissue (e.g., modeled as a power-law distribution) within the noise term of the statistical framework.

The model was shown to accurately predict the reduction in registration accuracy due to soft-tissue deformation in both 3D-2D and 3D-3D registration experiments. More importantly, it provided theoretical guidance for how to compensate for the effect of soft-tissue deformation — accurately predicting gradient-based similarity metrics to be robust to the effect of soft-tissue deformation.

Furthermore, in Chapter 4, the statistical framework was shown to provide important guidance in the development of a novel 3D-2D registration method for labeling vertebrae in radiographic images. The registration scenario presents multiple sources of content mismatch, including anatomical deformation and surgical instrumentation; however, by leveraging the

statistical model — utilizing sub-image cropping and gradient-based metrics to suppress these sources or noise — the registration method was able to effectively and robustly achieve accurate placement of the vertebral labels.

6.3 Aim 3: Investigate the Effect of Statistical Mismatch for CNN-Based Methods

In Chapter 5, CNN-based deformable registration was analyzed to understand how the statistical characteristics (e.g., noise, resolution, and deformation magnitude) of both the test data *and* the training data govern the resulting image registration accuracy. Experimental results in this study demonstrated the importance of statistical diversity in the training data in order to achieve a more generalizable CNN model — particularly showing the importance of ensuring that the test data is within the statistical range of the training data. For example, it was observed that increasing dose (and thus reducing noise) beyond that observed in the training data yielded no improvement in (and even reduced) registration accuracy. Such a point is particularly interesting in the context of the statistical framework in Aims 1 and 2, which would predict equal or better registration performance with the reduction of image noise.

Such observations on the generalizability of CNN performance with respect to training data statistics provide important guidance for the curation and augmentation of training data and emphasize the importance (and necessity) of training data diversity — since even for perfect ground truth data, generalizability of the model beyond the statistics observed in the input training data is limited.

The findings were further shown to be important in the development of an MR-CT synthesis, where three strategies for augmenting a training dataset were tested. Each strategy

further captured the statistical range of the test data, and each, respectively, yielded better synthesis performance. Interestingly, by including only one standard T1 MR image in the training dataset (combined with the other methods of augmentation), a significant boost in performance was observed when testing on standard T1 data — again showing that training data that encompasses the full statistical range (even if "book-ending" the range, rather than uniformly sampling it) may be sufficient when the full distribution of training data is not available.

6.4 Limitations and Future Work

Throughout this thesis, defining and quantifying the effect of factors that govern image registration error was shown to be an important step in obtaining knowledgeable guidance to the development of new registration methods. However, the models and methods used in this work come with limitations and opportunities for future work, briefly discussed below.

A major assumption taken in the statistical framework of Aims 1 and 2 is that the "true" signals between the two images are identical. The assumption is immediately violated for multi-modality image registration (i.e., images formed from fundamentally different contrast mechanisms) and likely necessitates an alternative framework that is specific to each modality. However, variations in the true signal are often present even within intra-modality registration. While it can be shown that the framework is insensitive to linear differences between the two images — or that some differences can be modeled as a noise source (e.g., deformation and content mismatch) — there are many scenarios in which a non-linear (but generally monotonic) relation exists between the two images — e.g., due to differences in kV in CT imaging or due to field inhomogeneity in MR. Such scenarios may still be categorized within the context of intra-modality image registration and can be solved using intra-modality similarity metrics.

Further investigation is warranted to determine how such differences may be incorporated in the framework (e.g., as a low-frequency noise) or if a new framework is needed.

Another limitation of the model pertains to the translation-only problem definition. Extension to rotation would yield insight on how errors in the geometric transformation compound (e.g., an error in rotation would likely lead to an increased error in translation); however, such a framework would likely necessitate a non-stationary representation of the image content — noting that salient structures far from the axis of rotation provide more rotational information than those close to the axis of rotation. Despite this limitation, however, the translation-only analysis still provides a useful framework for defining and modeling the factors that contribute to registration accuracy.

One of the most important factors to consider when developing a registration method is the robustness to local optima. While the presented statistical framework richly characterizes the region near the true solution (i.e., at the global optimum), it does not quantify the number or severity of local optima. As such, utilization of this framework should be paired with standard methods for achieving robustness to local optima. The most effective of these measures include proper initialization, multi-start optimizations, and multi-resolution pyramids. The third (multi-resolution pyramids) is particularly interesting as there is a possibility that the down-sampling and blur kernels could be set in an optimal manner according to the statistical framework of Chapter 2, leaving a potential opportunity for future work.

Obtaining the power spectrum representation of salient anatomy (or confounding deformable anatomy) remains a non-trivial step in the registration analysis pipeline. In the

experiments presented, we relied on noise and signal models presented in Tables 3.1 and 3.2 as well as multiple noise instantiations to obtain accurate power spectrum estimations. While such models can be built for each anatomical region and imaging system, an interesting avenue of research relates to the development of a reliable real-time estimation of the power spectra using the moving and fixed images alone (possibly in combination with cascaded systems analysis models). For example, one may imagine a registration method that alternates between updating the geometric transformation and estimating the power spectra, where the current registration estimate could be used to update the power spectrum estimates (e.g., image difference to achieve NPS estimation and image average to achieve signal power spectrum estimation). Therefore, each iteration would yield an improved statistical model for the guidance of post-processing parameters and similarity metric choice.

The extension of the statistical model to deformable registration using the locally rigid approximation (Sec. 5.2.2) provided an informative view (through the CRLB map) of the spatial dependence of registration accuracy. The use of deformation field regularization techniques, however, violated the assumption of an unbiased estimator and resulted in poor agreement between the model and experimental registration errors. Quantitatively modeling this bias would allow knowledgeable guidance on the type and strength of the regularization method.

CNN-based methods are difficult to characterize and predict performance due to their highly non-linear transfer characteristics, a broad variety of architectures, and a strong dependence on the training data. In Chapter 5, we examined three statistical characteristics related to the training data to understand the generalizability of the U-Net architecture. While the observations provided important insights regarding the diversity of the training data, there

is much work to be done to fully characterize CNN generalizability, particularly with respect to differences in model architecture, model size, and training loss function — all of which were fixed in the experiments of Chapter 5.

For any registration method to be translated into the clinical setting, it must be sufficiently accurate relative to the clinical objective, robust to gross registration failure, provide clinical benefit, and present minimal interruption to clinical workflow. While this dissertation has primarily focused on the first of these considerations (i.e., registration accuracy), the utility of this work relies heavily on ongoing translational research that addresses the latter factors — e.g., by designing robust optimization techniques, novel treatment techniques, and dedicated imaging and guidance systems. As such systems hold great promise for the advancement of interventional medicine, we hope that that the theoretical underpinnings established in this work provide a guide that will improve and accelerate the development of registration methods in clinical use.

Abbreviations

AWGN	Additive White Gaussian Noise
BFGS	Broyden-Fletcher-Goldfarb-Shanno
CBCT	Cone-Beam Computed Tomography
CC	Cross-Correlation
CMA-ES	Covariance Matrix Adaptation-Evolution Strategy
CNN	Convolutional Neural Network
CNR	Contrast-to-Noise Ratio
CRLB	Cramer-Rao Lower Bound
CT	Computed Tomography
DBS	Deep Brain Stimulation
DOF	Degree of Freedom
DQE	Detective Quantum Efficiency
DRR	Digitally Reconstructed Radiograph
ECC	Entropy Correlation Coefficient
ECT	Emission Computed Tomography
FFD	Free-Form Deformation
FIM	Fisher Information Matrix
FPD	Flat-Panel Detector
FWHM	Full Width at Half the Maximum
GAN	Generative Adversarial Network
GC	Gradient Correlation
GO	Gradient Orientation
GPU	Graphics Processing Unit
HU	Hounsfield Unit
IGRT	Image-Guided Radiotherapy
IQR	Inter-Quartile Range
IVD	Inter-Vertebral Distance
JMI	Joint-Histogram Mutual Information
MAE	Mean Absolute Error
MDCT	Multi-Detector Computed Tomography
MI	Mutual Information
MMI	Mattes Mutual Information
MR	Magnetic Resonance
MSD	Mean-Squared Difference
MSE	Mean-Squared Error
MTF	Modulation Transfer Function
NCC	Normalized Cross Correlation
NEQ	Noise-Equivalent Quanta
NMI	Normalized Mutual Information
NPS	Noise-Power Spectrum
NPWMF	Non-Prewhitening Matched Filter
PC	Phase Correlation
PDE	Projection Distance Error
PET	Positron Emission Tomography
PSF	Point-Spread Function
RF	Radio Frequency
RMSE	Root-Mean-Squared Error
ROI	Region of Interest
SDD	Source-to-Detector Distance
SDT	Statistical Decision Theory
SKE	Signal Known Exactly

SNR	Signal-to-Noise Ratio
SPECT	Single Photon Emission Tomography
SPR	Scatter-to-Primary Ratio
SR	Search Range
SRE	Statistical Registration Efficiency
SSD	Sum of Squared Difference
TCIA	The Cancer Imaging Archive
TDE	Time Delay Estimation
TRE	Target Registration Error
ZZB	Ziv-Zakai Bounds

Bibliography

[1] R. Park, T. Nyland, J. Lattimer, C. Miller, and J. Lebel, "B-mode gray-scale ultrasound: imaging artifacts and interpretation principles," *Vet. Radiol.*, vol. 22, no. 5, pp. 204–210, 1981.

[2] C. de Wiele, C. Lahorte, W. Oyen, O. Boerman, I. Goethals, G. Slegers, and R. A. Dierckx, "Nuclear medicine imaging to predict response to radiotherapy: a review," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 55, no. 1, pp. 5–15, 2003.

[3] C. Bremer, V. Ntziachristos, and R. Weissleder, "Optical-based molecular imaging: contrast agents and potential medical applications," *Eur. Radiol.*, vol. 13, no. 2, pp. 231–243, 2003.

[4] A. L. Vahrmeijer, M. Hutteman, J. R. Van Der Vorst, C. J. H. Van De Velde, and J. V Frangioni, "Image-guided cancer surgery using near-infrared fluorescence," *Nat. Rev. Clin. Oncol.*, vol. 10, no. 9, p. 507, 2013.

[5] J. H. Siewerdsen, I. A. Cunningham, and D. A. Jaffray, "A framework for noise-power spectrum analysis of multidimensional images," *Med. Phys.*, vol. 29, no. 11, pp. 2655–2671, 2002.

[6] D. J. Tward and J. H. Siewerdsen, "Cascaded systems analysis of the 3D noise transfer characteristics of flat-panel cone-beam CT," *Med. Phys.*, vol. 35, no. 12, p. 5510, 2008.

[7] M. F. Kijewski and P. F. Judy, "The noise power spectrum of CT images.," *Phys. Med. Biol.*, vol. 32, no. 5, pp. 565–575, 1987.

[8] D. A. Jaffray and J. H. Siewerdsen, "Cone-beam computed tomography with a flat-panel imager: initial performance characterization," *Med. Phys.*, vol. 27, no. 6, pp. 1311–1323, 2000.

[9] K. M. Hanson, "Detectability in computed tomographic images," *Med. Phys.*, vol. 6, no. 5, pp. 441–451, 1979.

[10] J. H. Siewerdsen, L. E. Antonuk, Y. El-Mohri, J. Yorkston, W. Huang, J. M. Boudry, and I. A. Cunningham, "Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (AMFPIs) for diagnostic radiology," *Med. Phys.*, vol. 24, no. 1, pp. 71–89, 1997.

[11] H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, "Objective assessment of image quality II Fisher information, Fourier crosstalk, and figures of merit for task performance," *J. Opt. Soc. Am. A*, vol. 12, no. 5, p. 834, 1995.

[12] ICRU, *ICRU Report 54 - Medical Imaging - The Assessment of Image Quality*. 1995.

[13] P. Prakash, W. Zbijewski, G. J. Gang, Y. Ding, J. W. Stayman, J. Yorkston, J. A.

Carrino, and J. H. Siewerdsen, “Task-based modeling and optimization of a cone-beam CT scanner for musculoskeletal imaging.,” *Med. Phys.*, vol. 38, no. 10, pp. 5612–5629, 2011.

[14] J. Xu, A. Sisniega, W. Zbijewski, H. Dang, J. W. Stayman, X. Wang, D. H. Foos, N. Aygun, V. E. Koliatsos, and J. H. Siewerdsen, “Modeling and design of a cone-beam CT head scanner using task-based imaging performance optimization,” *Phys. Med. Biol.*, vol. 61, no. 8, pp. 3180–3207, 2016.

[15] G. J. Gang, J. Siewerdsen, and J. Stayman, “Task-driven optimization of CT tube current modulation and regularization in model-based iterative reconstruction,” *Phys. Med. Biol.*, vol. 62, no. 12, p. 4777, 2017.

[16] S. Richard, D. B. Husarik, G. Yadava, S. N. Murphy, and E. Samei, “Towards task-based assessment of CT performance: system and object MTF across different reconstruction algorithms.,” *Med. Phys.*, vol. 39, no. 7, pp. 4115–4122, 2012.

[17] L. Yu, S. Leng, L. Chen, J. M. Kofler, R. E. Carter, and C. H. McCollough, “Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: impact of radiation dose and reconstruction algorithms.,” *Med. Phys.*, vol. 40, no. 4, p. 041908, 2013.

[18] M. J. Tapiovaara and R. F. Wagner, “SNR and noise measurements for medical imaging: I. A practical approach based on statistical decision theory,” *Phys. Med. Biol.*, vol. 38, no. 1, p. 71, 1993.

[19] A. E. Burgess, “Mammographic structure: data preparation and spatial statistics analysis,” in *Proc. SPIE*, 1999, vol. 3661. pp. 642–653.

[20] S. Richard, J. H. Siewerdsen, D. A. Jaffray, D. J. Moseley, and B. Bakhtiar, “Generalized DQE analysis of radiographic and dual-energy imaging using flat-panel detectors,” *Med. Phys.*, vol. 32, no. 5, pp. 1397–1413, 2005.

[21] L. Cockmartin, H. Bosmans, and N. W. Marshall, “Comparative power law analysis of structured breast phantom and patient images in digital mammography and breast tomosynthesis,” *Med. Phys.*, vol. 40, no. 8, p. 81920, 2013.

[22] L. Chen, C. K. Abbey, and J. M. Boone, “Association between power law coefficients of the anatomical noise power spectrum and lesion detectability in breast imaging modalities,” *Phys. Med. Biol.*, vol. 58, no. 6, p. 1663, 2013.

[23] M. Burger, J. Modersitzki, and L. Ruthotto, “A hyperelastic regularization energy for image registration,” *SIAM J. Sci. Comput.*, vol. 35, no. 1, pp. B132-B148, 2013.

[24] M.-C. Chiang, A. D. Leow, A. D. Klunder, R. A. Dutton, M. Barysheva, S. E. Rose, K. L. McMahon, G. I. De Zubicaray, A. W. Toga, and P. M. Thompson, “Fluid registration of diffusion tensor images using information theory,” *IEEE Trans. Med. Imaging*, vol. 27, no. 4, pp. 442–456, 2008.

[25] W. R. Crum, C. Tanner, and D. J. Hawkes, “Anisotropic multi-scale fluid registration:

evaluation in magnetic resonance breast imaging," *Phys. Med. Biol.*, vol. 50, no. 21, p. 5153, 2005.

[26] J.-P. Thirion, "Image matching as a diffusion process: an analogy with Maxwell's demons," *Med. Image Anal.*, vol. 2, no. 3, pp. 243–260, 1998.

[27] P. Cachier, E. Bardinet, D. Dormont, X. Pennec, and N. Ayache, "Iconic feature based nonrigid registration: the PASHA algorithm," *Comput. Vis. image Underst.*, vol. 89, no. 2–3, pp. 272–298, 2003.

[28] Y. Cao, M. I. Miller, R. L. Winslow, and L. Younes, "Large deformation diffeomorphic metric mapping of vector fields," *IEEE Trans. Med. Imaging*, vol. 24, no. 9, pp. 1216–1230, 2005.

[29] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, 2008.

[30] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61-S72, 2009.

[31] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Trans. Med. Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[32] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank, "Nonrigid multimodality image registration," in *Proc. SPIE*, 2001, vol. 4322. pp. 1609–1620.

[33] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, 1999.

[34] F. Maes, a Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information.," *IEEE Trans. Med. Imaging*, vol. 16, no. 2, pp. 187–98, Apr. 1997.

[35] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[36] A. Uneri, T. De Silva, J. Goerres, M. Jacobson, M. Ketcha, S. Reaungamornrat, G. Kleinszig, S. Vogt, A. Khanna, G. Osgood, J.-P. Wolinsky, and J. Siewerdsen, "Intraoperative evaluation of device placement in spine surgery using known-component 3D-2D image registration," *Phys. Med. Biol.*, vol. 62, no. 8, p. 3330, 2017.

[37] A. Uneri, A. S. Wang, Y. Otake, G. Kleinszig, S. Vogt, A. J. Khanna, G. L. Gallia, Z. L. Gokaslan, and J. H. Siewerdsen, "Evaluation of low-dose limits in 3D-2D rigid registration for surgical guidance," *Phys. Med. Biol*, vol. 59, no. 18, pp. 5329–5345, 2014.

[38] J. H. Siewerdsen, L. E. Antonuk, Y. El-Mohri, J. Yorkston, W. Huang, and I. A.

Cunningham, “Signal, noise power spectrum, and detective quantum efficiency of indirect-detection flat-panel imagers for diagnostic radiology,” *Med. Phys.*, vol. 25, no. 5, pp. 614–628, 1998.

[39] S. Richard and J. H. Siewerdsen, “Optimization of dual-energy imaging systems using generalized NEQ and imaging task,” *Med. Phys.*, vol. 34, no. 1, pp. 127–139, 2007.

[40] G. J. Gang, J. W. Stayman, W. Zbijewski, and J. H. Siewerdsen, “Task-based detectability in CT image reconstruction by filtered backprojection and penalized likelihood estimation.,” *Med. Phys.*, vol. 41, no. 8, p. 081902, 2014.

[41] G. J. Gang, J. Lee, J. W. Stayman, D. J. Tward, W. Zbijewski, J. L. Prince, and J. H. Siewerdsen, “Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance.,” *Med. Phys.*, vol. 38, no. 4, pp. 1754–1768, 2011.

[42] H. Dang, J. W. Stayman, J. Xu, W. Zbijewski, A. Sisniega, M. Mow, X. Wang, D. H. Foos, N. Aygun, V. E. Koliatsos, and others, “Task-based statistical image reconstruction for high-quality cone-beam CT,” *Phys. Med. Biol.*, vol. 62, no. 22, p. 8693, 2017.

[43] M. Ketcha, T. De Silva, R. Han, A. Uneri, J. Goerres, G. G, S. Vogt, and J. Siewerdsen, “Fundamental limits of image registration performance: Effects of image noise and resolution in CT-guided interventions,” in *Proc. SPIE*, 2017, vol. 10135, p. 1013508.

[44] M. D. Ketcha, T. De Silva, R. Han, A. Uneri, J. Goerres, M. W. Jacobson, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, “Effects of Image Quality on the Fundamental Limits of Image Registration Accuracy,” *IEEE Trans. Med. Imaging*, vol. 36, no. 10, pp. 1997–2009, 2017.

[45] I. S. Yetik and A. Nehorai, “Performance bounds on image registration,” *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1737–1749, 2006.

[46] T. Q. Pham, M. Bezuijen, L. J. Van Vliet, K. Schutte, and C. L. L. Hendriks, “Performance of optimal registration estimators,” in *Visual Information Processing XIV*, 2005, vol. 5817, pp. 133–145.

[47] M. L. Uss, B. Vozel, V. A. Dushepa, V. A. Komjak, and K. Chehdi, “A precise lower bound on image subpixel registration accuracy,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3333–3345, 2014.

[48] M. Xu, H. Chen, and P. K. Varshney, “Ziv-Zakai bounds on image registration,” *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1745–1755, 2009.

[49] C. Aguerrebere, M. Delbracio, A. Bartesaghi, and G. Sapiro, “Fundamental limits in multi-image alignment,” *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5707–5722, 2016.

[50] C. Aguerrebere, M. Delbracio, A. Bartesaghi, and G. Sapiro, “A Practical Guide to Multi-image Alignment,” *arXiv Prepr. arXiv1802.03280*, 2018.

[51] D. Robinson and P. Milanfar, “Fundamental Perfromance Limits in Image Registration,” *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1185–1199, 2004.

[52] C. Zhao, A. Carass, A. Jog, and J. L. Prince, “Effects of spatial resolution on image registration,” in *Proc. SPIE*, 2016, vol. 9784. p. 97840Y.

[53] A. R. Pineda, D. J. Tward, A. Gonzalez, and J. H. Siewerdsen, “Beyond noise power in 3D computed tomography: The local NPS and off-diagonal elements of the Fourier domain covariance matrix,” *Med. Phys.*, vol. 39, no. 6, p. 3240, 2012.

[54] W. J. I. Bangs, “Array Processing with Generalized Beamformers,” PhD dissertation, Yale University, New Haven, CT, 1971.

[55] A. Weiss and E. Weinstein, “Fundamental limitations in passive time delay estimation--Part I: Narrow-band systems,” *IEEE Trans. Acoust.*, vol. 31, no. 2, pp. 472–486, 1983.

[56] V. H. MacDonald and Peter M. Schultheiss, “Optimum Passive Bearing Estimation in a Spatially Incoherent Noise Environment,” *J. Acoust. Soc. Am.*, vol. 46, no. 1A, pp. 37–43, 1969.

[57] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 2002.

[58] C. H. Knapp and G. C. Carter, “The Generalized Correlation Method for Estimation of Time Delay,” *IEEE Trans. Acoust.*, vol. 24, no. 4, pp. 320–327, 1976.

[59] G. J. Gang, D. J. Tward, J. Lee, and J. H. Siewerdsen, “Anatomical background and generalized detectability in tomosynthesis and cone-beam CT,” *Med. Phys.*, vol. 37, no. 5, pp. 1948–1965, 2010.

[60] F. O. Bochud, J. F. Valley, F. R. Verdun, C. Hessler, and P. Schnyder, “Estimation of the noisy component of anatomical backgrounds.,” *Med. Phys.*, vol. 26, no. 7, pp. 1365–1370, 1999.

[61] J. Punnoose, J. Xu, A. Sisniega, W. Zbijewski, and J. H. Siewerdsen, “Technical Note: SPEKTR 3.0—A computational tool for x-ray spectrum modeling and analysis,” *Med. Phys.*, vol. 43, no. 8Part1, pp. 4711-4717, 2016.

[62] M. Unser, A. Aldroubi, and M. Eden, “B-spline signal processing. I. Theory,” *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 821–833, 1993.

[63] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, “The Design of SimpleITK.,” *Front. Neuroinform.*, vol. 7, p. 45, 2013.

[64] P. Thévenaz and M. Unser, “Optimization of mutual information for multiresolution image registration,” *IEEE Trans. Image Process*, vol. 9, no. 12, pp. 2083–2099, 2000.

[65] H. Foroosh, J. B. Zerubia, and M. Berthod, “Extension of phase correlation to subpixel registration,” *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 188–199, 2002.

[66] P. D. Welch, “The Use of Fast Fourier Transform for the Estimation of Power Spectra:

A Method Based on Time Averaging Over Short, Modified Periodograms," *IEEE Trans. Audio and Electroacoustics*, vol. 15. pp. 70–73, 1967.

[67] G. C. Carter, "Time delay estimation," Technical Report, *Nav. Underw. Syst. Cent.*, New London, CT, 1976.

[68] J. Baek and N. J. Pelc, "The noise power spectrum in CT with direct fan beam reconstruction.," *Med. Phys.*, vol. 37, no. 5, pp. 2074–2081, 2010.

[69] A. E. Burgess and P. F. Judy, "Signal detection in power-law noise: effect of spectrum exponents," *J. Opt. Soc. Am. A*, vol. 24, no. 12, pp. B52-B60, 2007.

[70] M. P. Eckstein, C. K. Abbey, and F. O. Bochud, "Visual signal detection in structured backgrounds. IV. Figures of merit for model performance in multiple-alternative forced-choice detection tasks with correlated responses," *J. Opt. Soc. Am. A*, vol. 17, no. 2, pp. 206–217, 2000.

[71] A. E. Burgess, "Visual signal detection with two-component noise: low-pass spectrum effects," *J. Opt. Soc. Am. A*, vol. 16, no. 3, pp. 694–704, 1999.

[72] S. Richard and J. H. Siewerdsen, "Optimization of dual-energy imaging systems using generalized NEQ and imaging task," *Med. Phys.*, vol. 34, no. 1, pp. 127–139, 2007.

[73] J. H. Siewerdsen and D. A. Jaffray, "Optimization of x-ray imaging geometry (with specific application to flat-panel cone-beam computed tomography)," *Med. Phys.*, vol. 27, no. 8, p. 1903, 2000.

[74] G. J. Gang, J. H. Siewerdsen, and J. W. Stayman, "Task-Driven Optimization of Fluence Field and Regularization for Model-Based Iterative Reconstruction in Computed Tomography," *IEEE Trans. Med. Imaging*, vol. 36, no. 12, pp. 2424–2435, 2017.

[75] A. E. Burgess, "Statistically defined backgrounds: performance of a modified nonprewhitening observer model," *J. Opt. Soc. Am. A*, vol. 11, no. 4, pp. 1237–1242, 1994.

[76] J. P. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," *J. Opt. Soc. Am. A*, vol. 9, no. 5, pp. 649–658, 1992.

[77] M. D. Ketcha, T. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, "A Statistical Model for Rigid Image Registration Performance : The Influence of Soft-Tissue Deformation as a Confounding Noise Source," *IEEE Trans. Med. Imaging*, vol. 38. no. 9, pp. 2016-2027, 2019.

[78] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and others, "A comparison of similarity measures for use in 2-D-3-D medical image registration," *IEEE Trans. Med. Imaging*, vol. 17, no. 4, pp. 586–595, 1998.

[79] Y. Otake, S. Schafer, J. W. Stayman, W. Zbijewski, G. Kleinszig, R. Graumann, a J. Khanna, and J. H. Siewerdsen, "Automatic localization of vertebral levels in x-ray fluoroscopy using 3D-2D registration: a tool to reduce wrong-site surgery.," *Phys. Med.*

Biol., vol. 57, no. 17, pp. 5485–508, 2012.

[80] T. De Silva, A. Uneri, M. D. Ketcha, S. Reaungamornrat, G. Kleinszig, S. Vogt, N. Aygun, S. F. Lo, J. P. Wolinsky, and J. H. Siewerdsen, "3D-2D image registration for target localization in spine surgery: investigation of similarity metrics providing robustness to content mismatch," *Phys. Med. Biol.*, vol. 61, no. 8, p. 3009, 2016.

[81] J. H. Siewerdsen, "Signal, noise, and detective quantum efficiency of amorphous-silicon:hydrogen flat-panel imagers," PhD dissertation, University of Michigan, Ann Arbor, MI, 1998.

[82] I. Reiser, S. Lee, and R. M. Nishikawa, "On the orientation of mammographic structure," *Med. Phys.*, vol. 38, no. 10, pp. 5303–5306, 2011.

[83] M. G. Mody, A. Nourbakhsh, D. L. Stahl, M. Gibbs, M. Alfawareh, and K. J. Garges, "The prevalence of wrong level surgery among spine surgeons.," *Spine*, vol. 33, no. 2, pp. 194–198, 2008.

[84] Y. Otake, A. S. Wang, J. Webster Stayman, A. Uneri, G. Kleinszig, S. Vogt, a J. Khanna, Z. L. Gokaslan, and J. H. Siewerdsen, "Robust 3D-2D image registration: application to spine interventions and vertebral labeling in the presence of anatomical deformation.," *Phys. Med. Biol.*, vol. 58, no. 23, pp. 8535–8553, 2013.

[85] T. De Silva, S.-F. L. Lo, N. Aygun, D. M. Aghion, A. Boah, R. Petteys, A. Uneri, M. D. Ketcha, T. Yi, S. Vogt, and others, "Utility of the LevelCheck algorithm for decision support in vertebral localization," *Spine*, vol. 41, no. 20, p. E1249, 2016.

[86] M. D. Ketcha, T. De Silva, A. Uneri, M. W. Jacobson, J. Goerres, G. Kleinszig, S. Vogt, J. P. Wolinsky, and J. H. Siewerdsen, "Multi-stage 3D-2D registration for correction of anatomical deformation in image-guided spine surgery," *Phys. Med. Biol.*, vol. 62, no. 11, p. 4604, 2017.

[87] S. Benameur, M. Mignotte, S. Parent, H. Labelle, W. Skalli, and J. de Guise, "3D/2D registration and segmentation of scoliotic vertebrae using statistical models," *Comput. Med. Imaging Graph.*, vol. 27, no. 5, pp. 321–337, 2003.

[88] M. Prümmer, J. Hornegger, M. Pfister, and A. Dörfler, "Multi-modal 2D-3D non-rigid registration," in *Proc. SPIE*, 2006, vol. 6144, pp. 297–308.

[89] D. Rivest-Hénault, H. Sundar, and M. Cheriet, "Nonrigid 2D/3D registration of coronary artery models with live fluoroscopy for guidance of cardiac interventions.," *IEEE Trans. Med. Imaging*, vol. 31, no. 8, pp. 1557–1572, 2012.

[90] A. Guyot, A. Varnavas, T. Carrell, and G. Penney, "Non-rigid 2D-3D registration using anisotropic error ellipsoids to account for projection uncertainties during aortic surgery.," in *Proc. MICCAI*, 2013, pp. 179–186.

[91] A. Uneri, J. Goerres, T. De Silva, M. W. Jacobson, M. D. Ketcha, S. Reaungamornrat, G. Kleinszig, S. Vogt, A. J. Khanna, J.-P. Wolinsky, and others, "Deformable 3D-2D registration of known components for image guidance in spine surgery," in *Proc.*

MICCAI, 2016, pp. 124–132.

[92] M. Groher, "2D-3D Registration of Vascular Images," PhD dissertation, Technische Universität München, Munich, Germany, 2008

[93] G. P. Penney, J. A. Little, J. Weese, D. L. G. Hill, and D. J. Hawkes, "Deforming a preoperative volume to represent the intraoperative scene.," *Comput. Aided Surg.*, vol. 7, no. 2, pp. 63–73, 2002.

[94] J. Schmid and C. Chênes, "Segmentation of X-ray images by 3D-2D registration based on multibody physics," in *Asian Conference on Computer Vision*, 2014, pp. 674–687.

[95] G. Penney, "Registration of tomographic images to X-ray projections for use in image guided interventions," PhD dissertation, University of London, London, England, 2000.

[96] J. Weese, G. P. Penney, P. Desmedt, T. M. Buzug, D. L. G. Hill, and D. J. Hawkes, "Voxel-based 2-D/3-D registration of fluoroscopy images and CT scans for image-guided surgery," *IEEE Trans. Inf. Technol. Biomed.*, vol. 1, no. 4, pp. 284–293, 1997.

[97] S. Ourselin, A. Roche, S. Prima, and N. Ayache, "Block Matching: A General Framework to Improve Robustness of Rigid Registration of Medical Images," in *Proc. MICCAI*, 2000, pp. 557–566.

[98] Shan Zhu and Kai-Kuang Ma, "A new diamond search algorithm for fast block-matching motion estimation," *IEEE Trans. Image Process.*, vol. 9, no. 2, pp. 287–290, 2000.

[99] A. Varnavas, T. Carrell, and G. Penney, "Fully automated 2D-3D registration and verification.," *Med. Image Anal.*, vol. 26, no. 1, pp. 108–119, 2015.

[100] J. E. Scholtz, J. L. Wichmann, M. Kaup, S. Fischer, J. M. Kerl, T. Lehnert, T. J. Vogl, and R. W. Bauer, "First performance evaluation of software for automatic segmentation, labeling and reformation of anatomical aligned axial images of the thoracolumbar spine at CT," *Eur. J. Radiol.*, vol. 84, no. 3, pp. 437–442, 2015.

[101] T. De Silva, A. Uneri, M. D. Ketcha, S. Reaungamornrat, J. Goerres, M. W. Jacobson, S. Vogt, G. Kleinszig, A. J. Khanna, J. P. Wolinsky, and others, "Registration of MRI to intraoperative radiographs for target localization in spinal interventions," *Phys. Med. Biol.*, vol. 62, no. 2, p. 684, 2017.

[102] M. D. Ketcha, T. De Silva, A. Uneri, G. Kleinszig, S. Vogt, J.-P. Wolinsky, and J. H. Siewerdsen, "Automatic masking for robust 3D-2D image registration in image-guided spine surgery," in *Proc. SPIE*, 2016, vol. 9786, p. 97860A.

[103] B. Cabral, N. Cam, and J. Foran, "Accelerated Volume Rendering and Tomographic Reconstruction Using Texture Mapping Hardware," in *Proceedings of the 1994 Symposium on Volume Visualization*, 1994, pp. 91–98.

[104] N. Hansen, "The CMA Evolution Strategy : A Comparing Review," In *Towards a New Evolutionary Computation,* Springer, Berlin, Heidelberg, pp. 75–102, 2006.

[105] J. L. Bentley, “Multidimensional binary search trees used for associative searching,” *Commun. ACM*, vol. 18, no. 9, pp. 509–517, 1975.

[106] F. L. Markley, Y. Cheng, J. L. Crassidis, and Y. Oshman, “Averaging quaternions,” *J. Guid. Control. Dyn.*, vol. 30, no. 4, pp. 1193–1197, 2007.

[107] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, “User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability.,” *NeuroImage*, vol. 31, no. 3, pp. 1116–1128, 2006.

[108] S.-F. L. Lo, Y. Otake, V. Puvanesarajah, A. S. Wang, A. Uneri, T. De Silva, S. Vogt, G. Kleinszig, B. D. Elder, C. R. Goodwin, T. A. Kosztowski, J. A. Liauw, M. Groves, A. Bydon, D. M. Sciubba, T. F. Witham, J.-P. Wolinsky, N. Aygun, Z. L. Gokaslan, and J. H. Siewerdsen, “Automatic localization of target vertebrae in spine surgery: clinical evaluation of the LevelCheck registration algorithm.,” *Spine*, vol. 40, no. 8, pp. E476-83, 2015.

[109] J. Fan, X. Cao, Q. Wang, P.-T. Yap, and D. Shen, “Adversarial Learning for Mono-or Multi-Modal Registration,” *Med. Image Anal.*, vol. 58, p. 101545, 2019.

[110] C. Wang, G. Papanastasiou, A. Chartsias, G. Jacenkow, S. A. Tsaftaris, and H. Zhang, “FIRE: Unsupervised bi-directional inter-modality registration using deep networks,” *arXiv Prepr. arXiv1907.05062*, 2019.

[111] H. Uzunova, M. Wilms, H. Handels, and J. Ehrhardt, “Training CNNs for image registration from few samples with model-based data augmentation,” in *Proc. MICCAI*, 2017, pp. 223–231.

[112] K. A. J. Eppenhof and J. P. W. Pluim, “Pulmonary CT Registration through Supervised Learning with Convolutional Neural Networks,” *IEEE Trans. Med. Imaging*, vol. 38, no. 5, pp. 1097–1105, 2018.

[113] M. Jaderberg, K. Simonyan, A. Zisserman, and others, “Spatial transformer networks,” in *Advances in Neural Information Processing Systems*, 2015, pp. 2017–2025.

[114] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, “End-to-end unsupervised deformable image registration with a convolutional neural network,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2017, pp. 204–212.

[115] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, “Nonrigid registration using free-form deformations: application to breast MR images,” *IEEE Trans. Med. Imaging*, vol. 18, no. 8, pp. 712–721, 1999.

[116] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, “Symmetric log-domain diffeomorphic registration: A demons-based approach,” in *Proc. MICCAI*, 2008, pp. 754–761.

[117] R. Castillo, E. Castillo, R. Guerra, V. E. Johnson, T. McPhail, A. K. Garg, and T. Guerrero, “A framework for evaluation of deformable image registration spatial

accuracy using large landmark point sets," *Phys. Med. Biol.*, vol. 54, no. 7, p. 1849, 2009.

[118] E. Castillo, R. Castillo, J. Martinez, M. Shenoy, and T. Guerrero, "Four-dimensional deformable image registration using trajectory modeling," *Phys. Med. Biol.*, vol. 55, no. 1, p. 305, 2009.

[119] J. Vandemeulebroucke, S. Rit, J. Kybic, P. Clarysse, and D. Sarrut, "Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs," *Med. Phys.*, vol. 38, no. 1, pp. 166–178, 2011.

[120] J. Vandemeulebroucke, D. Sarrut, P. Clarysse, and others, "The POPI-model, a point-validated pixel-based breathing thorax model," in *XVth International Conference on the Use of Computers in Radiation Therapy (ICCR)*, 2007, vol. 2, pp. 195–199.

[121] M. D. Ketcha, T. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, "Effect of statistical mismatch between training and test images for CNN-based deformable registration," in *Proc. SPIE*, 2019, vol. 10949, p. 109490T.

[122] M. D. Ketcha, T. S. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, "Learning-based deformable image registration: effect of statistical mismatch between train and test images," *J. Med. Imaging*, vol. 6, no. 4, p. 44008, 2019.

[123] M.-M. Rohé, M. Datar, T. Heimann, M. Sermesant, and X. Pennec, "SVF-Net: Learning Deformable Image Registration Using Shape Matching," in *Proc. MICCAI*, 2017, pp. 266–274.

[124] H. Sokooti, B. de Vos, F. Berendsen, B. P. F. Lelieveldt, I. Išgum, and M. Staring, "Nonrigid image registration using multi-scale 3D convolutional neural networks," in *Proc. MICCAI*, 2017, pp. 232–239.

[125] Y. Hu, M. Modat, E. Gibson, W. Li, N. Ghavami, E. Bonmati, G. Wang, S. Bandula, C. M. Moore, M. Emberton, and others, "Weakly-supervised convolutional neural networks for multimodal image registration," *Med. Image Anal.*, vol. 49, pp. 1–13, 2018.

[126] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.

[127] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1125–1134.

[128] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network.," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4681–4690.

[129] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International*

Conference on Computer Vision, 2017, pp. 2223–2232.

[130] H. Yang, J. Sun, A. Carass, C. Zhao, J. Lee, Z. Xu, and J. Prince, "Unpaired brain MR-to-CT synthesis using a structure-constrained CycleGAN," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2018, pp. 174–182.

[131] Y. Hiasa, Y. Otake, M. Takao, T. Matsuoka, K. Takashima, A. Carass, J. L. Prince, N. Sugano, and Y. Sato, "Cross-modality image synthesis from unpaired data using CycleGAN," in *International Workshop on Simulation and Synthesis in Medical Imaging*, 2018, pp. 31–41.

[132] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. MICCAI,* 2015, pp. 234–241.

[133] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *CoRR*, vol. abs/1412.6, 2014.

[134] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic Demons Using ITK's Finite Difference Solver Hierarchy," *The Insight Journal*, pp. 1–8, 2007.

[135] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, and others, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *J. Digit. Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[136] H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M. Singh, and M.-H. Yang, "Diverse image-to-image translation via disentangled representations," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 35–51.

Curriculum Vitae

The Johns Hopkins University School of Medicine

Michael Daniel Ketcha — Date of this Version: 7/22/2020

Educational History:

Ph.D. *Biomedical Engineering* — 2020 (Expected)
Johns Hopkins University
Mentor: Jeffrey H. Siewerdsen, Ph.D.

B.S. *Biomedical Engineering* — 2014
Johns Hopkins University
Minor: Computer Integrated Surgery

Other Professional Experience:

Research Intern	Summer 2019	Medtronic, Littleton, MA
Image Scientist	2013-2015	Sonavex, Inc., Baltimore, MD
Undergrad Researcher	2013-2014	Center for Imaging Science, Baltimore, MD

Honors and Awards:

2019	Siebel Scholar	Seibel Foundation
2017	Young Scientist Award	SPIE Medical Imaging
2017	Conference Finalist of the Robert F. Wagner Paper Award	SPIE Medical Imaging
2016	Expanding Horizons Travel Grant	AAPM

Peer-Reviewed Publications:

[1] R. Han, A. Uneri, **M.D. Ketcha**, R. Vijayan, N. Sheth, P. Wu, P. Vagdargi, S. Vogt, G. Kleinszig, G.M. Osgood, and J.H. Siewerdsen. Multi-body 3D-2D registration for image-guided reduction of pelvic dislocation in orthopaedic trauma surgery. *Phys. Med. Biol., (in press) 2020.*

[2] T.S. De Silva, S.S. Vedula, A. Perdomo-Pantoja, R.C. Vijayan, S.A. Doerr, A. Uneri, R. Han, **M.D. Ketcha**, R.L. Skolasky, T. Witham, N. Theodore, and J.H. Siewerdsen. SpineCloud: image analytics for predictive modeling of spine surgery outcomes. *J. Med. Imaging*, vol. *7, no.* 3, p.031502, 2020.

[3] N.M. Sheth, T. De Silva, A. Uneri, **M.D. Ketcha**, R. Han, R. Vijayan, G.M. Osgood, and J.H. Siewerdsen, 2019. A Mobile Isocentric C-Arm for Intraoperative Cone-Beam

CT: Technical Assessment of Dose and 3D Imaging Performance. *Med. Phys.* Vol. 47, no.3, p. 958-974, 2020.

[4] **M.D. Ketcha**, T.S. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J.H. Siewerdsen. Learning-based deformable image registration: effect of statistical mismatch between train and test images. *J. Med. Imaging*, vol. *6, no.* 4, p.044008, 2019.

[5] R. Vijayan, T. De Silva, R. Han, X. Zhang, A. Uneri, S. Doerr, **M.D. Ketcha**, A. Perdomo-Pantoja, N. Theodore, and J.H. Siewerdsen. Automatic pedicle screw planning using atlas-based registration of anatomy and reference trajectories. *Phys. Med. Biol.*, vol. *64, no.* 16, p.165020, 2019.

[6] R. Han, A. Uneri, T. De Silva, **M. D. Ketcha**, J. Goerres, S. Vogt, G. Kleinszig,G. Osgood, and J. Siewerdsen, "Atlas-based automatic planning and 3d–2d fluoroscopic guidance in pelvic trauma surgery," *Phys. Med. Biol.*, vol. 64, no. 9, p. 095022, 2019.

[7] **M. D. Ketcha,** T. De Silva, R. Han, A. Uneri, J. Goerres, M. Jacobson, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, "A Statistical Model for Rigid Image Registration Performance: The Influence of Soft-Tissue Deformation as a Confounding Noise Source," *IEEE Trans. Med. Imaging,* vol. 38, no. 9, p. 2016, 2019.

[8] M. W. Jacobson, **M. D. Ketcha**, S. Capostagno, A. Martin, A. Uneri, J. Goerres, T. De Silva, S. Reaungamornrat, R. Han, A. Manbachi, *et al*., "A line fiducial method for geometric calibration of cone-beam CT systems with diverse scan trajectories," *Phys. Med. Biol.*, vol. 63, no. 2, p. 25030, 2018.

[9] T. De Silva, J. Punnoose, A. Uneri, M. Mahesh, J. Goerres, M. Jacobson, **M. D. Ketcha**, A. Manbachi, S. Vogt, G. Kleinszig, *et al*., "Virtual fluoroscopy for intraoperative C-arm positioning and radiation dose reduction," *J. Med. Imaging*, vol. 5, no. 1, p. 15005, 2018.

[10] A. Manbachi, T. De Silva, A. Uneri, M. Jacobson, J. Goerres, **M. D. Ketcha**, R. Han, N. Aygun, D. Thompson, X. Ye, and others, "Clinical translation of the levelcheck decision support algorithm for target localization in spine surgery," *Ann. Biomed. Eng.*, vol. 46, no. 10, pp. 1548–1557, 2018.

[11] R. Han, T. De Silva, **M. D. Ketcha**, A. Uneri, and J. H. Siewerdsen, "A momentum-based diffeomorphic demons framework for deformable MR-CT image registration," *Phys. Med. Biol.*, vol. 63, no. 21, p. 215006, 2018.

[12] T. De Silva, A. Uneri, X. Zhang, **M. D. Ketcha**, R. Han, N. Sheth, A. Martin, S. Vogt, G. Kleinszig, A. Belzberg, and others, "Real-time, image-based slice-to-volume registration for ultrasound-guided spinal intervention," *Phys. Med. Biol.*, vol. 63, no. 21, p. 215016, 2018.

[13] R. Tang, **M. D. Ketcha**, A. Badea, E. D. Calabrese, D. S. Margulies, J. T. Vogelstein, C. E. Priebe, and D. L. Sussman, "Connectome Smoothing via Low-rank Approximations," *IEEE Trans. Med. Imaging*, vol. 38, no. 6, p. 1446-1456, 2018.

[14] **M. D. Ketcha**, T. De Silva, A. Uneri, M. W. Jacobson, J. Goerres, G. Kleinszig, S. Vogt, J. P. Wolinsky, and J. H. Siewerdsen, "Multi-stage 3D-2D registration for correction of anatomical deformation in image-guided spine surgery," *Phys. Med. Biol.*, vol. 62, no. 11, p. 4604, 2017.

[15] J. Goerres, A. Uneri, T. De Silva, **M. D. Ketcha**, S. Reaungamornrat, M. Jacobson, S. Vogt, G. Kleinszig, G. Osgood, J. P. Wolinsky, and others, "Spinal pedicle screw planning using deformable atlas registration," *Phys. Med. Biol.*, vol. 62, no. 7, p. 2871, 2017.

[16] T. De Silva, A. Uneri, **M. D. Ketcha**, S. Reaungamornrat, J. Goerres, M. W. Jacobson, S. Vogt, G. Kleinszig, A. J. Khanna, J. P. Wolinsky, and others, "Registration of MRI to intraoperative radiographs for target localization in spinal interventions," *Phys. Med. Biol.*, vol. 62, no. 2, p. 684, 2017.

[17] A. Uneri, T. De Silva, J. Goerres, M. W. Jacobson, **M. D. Ketcha**, S. Reaungamornrat, G. Kleinszig, S. Vogt, A. J. Khanna, G. M. Osgood, and others, "Intraoperative evaluation of device placement in spine surgery using known-component 3D--2D image registration," *Phys. Med. Biol.*, vol. 62, no. 8, p. 3330, 2017.

[18] A. Manbachi, T. De Silva, A. Uneri, M. W. Jacobson, J. Goerres, **M. D. Ketcha**, R. Han, N. Aygun, D. A. Thompson, X. Ye, and others, "Clinical Translation of the LevelCheck Algorithm for Automatic Localization of Target Vertebrae in Spine Surgery," *Spine J.*, vol. 17, no. 10, p. S202, 2017.

[19] J. Goerres, A. Uneri, M. Jacobson, B. Ramsay, T. De Silva, **M. D. Ketcha**, R. Han, A. Manbachi, S. Vogt, G. Kleinszig, and others, "Planning, guidance, and quality assurance of pelvic screw placement using deformable image registration," *Phys. Med. Biol.*, vol. 62, no. 23, p. 9018, 2017.

[20] **M. D. Ketcha**, T. De Silva, R. Han, A. Uneri, J. Goerres, M. Jacobson, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, "Effects of Image Quality on the Fundamental Limits of Image Registration Accuracy," *IEEE Trans. Med. Imaging*, vol. 0062, no. c, pp. 1–1, 2017.

[21] T. De Silva, A. Uneri, **M. D. Ketcha**, S. Reaungamornrat, G. Kleinszig, S. Vogt, N. Aygun, S. F. Lo, J. P. Wolinsky, and J. H. Siewerdsen, "3D--2D image registration for target localization in spine surgery: investigation of similarity metrics providing robustness to content mismatch," *Phys. Med. Biol.*, vol. 61, no. 8, p. 3009, 2016.

[22] T. De Silva, S.-F. L. Lo, N. Aygun, D. M. Aghion, A. Boah, R. Petteys, A. Uneri, **M. D. Ketcha**, T. Yi, S. Vogt, and others, "Utility of the LevelCheck algorithm for decision support in vertebral localization," *Spine (Phila. Pa. 1976).*, vol. 41, no. 20, p. E1249, 2016.

[23] S. Reaungamornrat, T. De Silva, A. Uneri, J. Goerres, M. Jacobson, **M. D. Ketcha**, S. Vogt, G. Kleinszig, A. J. Khanna, J. P. Wolinsky, and others, "Performance evaluation of MIND demons deformable registration of MR and CT images in spinal interventions," *Phys. Med. Biol.*, vol. 61, no. 23, p. 8276, 2016.

[24] M. I. Miller, J. T. Ratnanather, D. J. Tward, T. Brown, D. S. Lee, **M. D. Ketcha**, K. Mori, M.-C. Wang, S. Mori, M. S. Albert, and others, "Network neurodegeneration in Alzheimer's disease via MRI based shape diffeomorphometry and high-field atlasing," *Front. Bioeng. Biotechnol.*, vol. 3, p. 54, 2015.

Conference Presentations:

[1] **M. D. Ketcha**, C.K. Jones, P. Wu, R. Han, A. Uneri, J. Lee, M. Luciano, W.S. Anderson, and J.H. Siewerdsen. "Deformable MR to Cone-Beam CT Registration for High-Precision Neuro-Endoscopic Surgery." National Image Guided Therapy Workshop 2020. Virtual Conference, Poster Presentation (April 2020).

[2] **M. D. Ketcha**, T. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J. Siewerdsen, "A Statistical Model Relating Image Quality to Image Registration Accuracy in Image-Guided Surgery," *Bull. Am. Phys. Soc.*, Boston, MA Oral Presentation (2019).

[3] **M. D. Ketcha**, T. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J. Siewerdsen, J.H., "Effect of statistical mismatch between training and test images for CNN-based deformable registration." *SPIE Medical Imaging*, San Diego, CA, Oral Presentation (February 2019)

[4] **M. D. Ketcha**, T. De Silva, R. Han, A. Uneri, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, J.H., "A statistical model for image registration performance: effect of tissue deformation." *SPIE Medical Imaging*, Houston, TX, Oral Presentation (February 2018)

[5] **M. D. Ketcha**, T. de Silva, R. Han, A. Uneri, J. Goerres, M. Jacobson, S. Vogt, G. Kleinszig, and J. H. Siewerdsen, "Fundamental limits of image registration performance: effects of image noise and resolution in CT-guided interventions," *SPIE Medical Imaging*, Orlando, FL, Oral Presentation (February 2017)

[6] **M. D. Ketcha**, T. De Silva, A. Uneri, G. Kleinszig, S. Vogt, J.-P. Wolinsky, and J. H. Siewerdsen, "Automatic masking for robust 3D-2D image registration in image-guided spine surgery," *SPIE Medical Imaging*, San Diego, CA, Oral Presentation (February 2016)

Patent Applications:

[1] J. H. Siewerdsen, M. W. Jacobson, and **M. D. Ketcha.** "Geometric calibration for cone beam ct using line fiducials." March 17, 2020. U.S. Patent App. 16/494,439

[2] **M. D. Ketcha**, W. T. De Silva, A. Uneri, J.-P. Wolinsky, and J. H. Siewerdsen, "Method for deformable 3d-2d registration using multiple locally rigid registrations," Apr. 16 2019. US Patent App. 10/262,424.

[3] J. H. Siewerdsen, W. T. De Silva, A. Uneri, **M. D. Ketcha**, S. Reaungamornrat, and J.-P. Wolinsky, "MR-LevelCheck-2: method for localization of structures in projection images." (Filed 2017).

Service and Leadership:

2018-2019	**Co-Director of Internships (Co-Director of Careers, 2015-2016):** *Biomedical Engineering Extramural Development in Graduate Education (BME EDGE).*
2017-2020	**Co-Founder of Hopkins Hikers**
2017	**BCI-EDGE Student Advisory Committee**
2016	**Teaching Assistant,** *Statistical Mechanics and Thermodynamics*
2015	**Co-Director 2015 Hopkins Imaging Conference**
2014	**Learning Den Teaching Assistant,** *Systems and Controls*

2012-2014 **Alpha Phi Omega Community Service Fraternity**

2010-2014 **Instructor for Johns Hopkins Office of Experiential Education**

Reviewer:

IEEE Transactions on Medical Imaging
IEEE Transactions on Biomedical Engineering
Medical Image Analysis
Computers in Biology and Medicine